# COMMENTS BY COUNTRIES ON THE DISCUSSION PAPER OF THE TASK FORCE ON CONFIDENTIALITY AND MICRODATA

## Table of Contents



***

This document collates the comments sent by countries and international organizations to the discussion paper of the Task Force on Confidentiality and Microdata (CES/2004/WP.1). The Task Force and the ECE secretariat thank all persons/organizations who contributed to the discussion. The different views expressed provide valuable input for preparing the final version of the paper that could then be endorsed by the CES and its Bureau.

The document begins with general comments to the whole discussion paper. After that, the comments are given in the order following the structure of the discussion paper. When comments refer to specific paras, the original text is provided in italics (in dark blue colour) for easy reference. In order to keep the statements on a certain topic in one place, the comments/reactions to the questions in paras 69-75 of the discussion paper are referred to the parts of text where those issues are considered in more detail.

## I GENERAL COMMENTS

EUROSTAT:

Eurostat supports the general approach of the paper that aims to resolve the tension between researcher community and producers of official statistics (mainly national statistical organisations - NSOs).

UNITED STATES:

This report has comprehensively documented many of the important areas that need to be addressed to allow data to be shared among countries in a safe, confidential manner. It correctly points out that some countries are much further along in this effort than others. While the report includes recommendations that are reasonable, there may be potential conflicts between statutory provisions for privacy/confidentiality and the final principles.

ECB

No comment.

NETHERLANDS:

At your request Statistics Netherlands (CBS) approached a number of Dutch instances to comment on the CES paper on confidentiality and microdata. The Netherlands Privacy Authority (CBP) did not feel the need to comment. It has always actively, critically and constructively, supported Statistics Netherlands in our microdata release policy and legislation.

All instances and persons consulted agree with the essence of the analysis and recommendations. There is ample support for substantive accessibility of official statistical microdata in the interest of research and society.

Let me quote from the consecutive comments:

"The paper offers a good overview of problems."

"I agree with the ideas and concept in this paper, and the Issues for Discussion."

"The recommendations are well worth pursuing."

"Share responsibility for ensuring maximum return on investments in data collection ... maintain the trust [and privacy] of respondents also is a shared responsibility."

There is a general preference for the pragmatic cost-benefit approach of the issues involved which is called risk management in your paper. On the issue of the costs of microdata the opinions do not fully converge. The continued support of the Privacy Commissions is deemed essential. Amongst the academic research community itself there is an increasing awareness of privacy issues and an active policy to draft and maintain codes of conduct. This, by the way, is in line with the policy of the Netherlands Privacy Authority, to look for self-regulation where possible and appropriate.

NETHERLANDS [Stoop]:

This is a very useful paper giving guidance in how statistical microdata can be made available for research purposes while safeguarding confidentiality up to a certain level and while minimizing the risk of identification of individual units. The recommendations are definitely worth pursuing.

NETHERLANDS [den Butter]:

The paper offers a good overview of problems associated with the use of statistical data, micro data in particular, of NSO's by members of the scientific community

The paper does not consider the use of administrative data and the resulting micro databases by individual analysts.

NETHERLANDS [Vollering]:

The collection of microdata for scientific research purposes is an expensive enterprise. The government (national, provincial, local, etc.), non-profit organizations (like National Statistical Offices/ NSOs and national research councils) and scientific research organizations (like universities), spending much money on data collection, share responsibility for ensuring maximum return on investments in data collection. This implies that microdata is used as intensively as possible. On the other hand, these organizations must maintain the trust of respondents. The privacy of the respondents must be guaranteed. This also is a shared responsibility of the government, non-profit organizations and scientific research organizations.

AUSTRIA:

Context of Austria's Statements:
With regard to the discussion paper it must be explained that a circulation of personal statistical data for scientific purposes is basically not possible on the basis of the Austrian Federal Statistics Act. Pursuant to article 31, access to individual data made anonymous for scientific purposes is possible only on certain restricted conditions. Moreover, we should like to point out that EU Regulation no.831/2002 governs the access to confidential data for scientific purposes very strictly. Our statement must be regarded in this context.

## II BACKGROUND

UNITED STATES:

It is not clear from the introduction that the paper is concerned with developing principles under which NSOs should disseminate microdata that would encourage consistency across countries and thereby facilitate international use of microdata. This should be made clear.

## III A FRAMEWORK TO SUPPORT DISCUSSION

UNITED STATES:

Section 2 has omitted one possible statistical product which might help: synthetic microdata files that mimic to a great extent the correlations inherent in the original microdata (though there is a passing reference to this approach later in the paper).

The definition of "microdata" appears on page 3 under the Data Laboratories dissemination stream Notes column; we suggest that it should appear as a term in the left column with its definition in the right column.

NETHERLANDS [Stoop]:

Under 7A. anonymisation is clearly more than the removal of name and address, which is at variance with point 9.

## IV THE PERSPECTIVE OF THE NATIONAL STATISTICAL OFFICE

**[RESPONDENT TRUST]**

*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.*

EUROSTAT:

Moreover, there is a necessity of a more harmonised approach to data protection for the following reasons. First, the so-called "mad cow" effect (a crisis in one country could affect other countries) calls for levels of protection to be high everywhere (if a national system is weak, this threads the credibility of the protection implemented in its neighbours). Second, some countries refuse to disclose data, even in conditions which will protect confidentiality if other countries have a different policy; the reasoning behind is that they refuse to give full visibility on their economy if the others do not (this argument has been used for business data).

**[QUALITY]**

*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.*

NETHERLANDS [Kooiman]:

An additional concern sometimes encountered is that researchers might draw invalid inferences from the survey data, due to ignorance about the meaning of the data, sampling issues, questionnaire, editing procedures, imputations and so on. In part this can be countered by preparing good documentation.

NETHERLANDS [Dekker]:

Data quality: If the quality of NSO's data is poor, there will be no demand by the research community (it's a market or self-regulating mechanism.
Perhaps even more important for secondary use is the quality of documentation (metadata). To improve (and be able to check or benchmark) quality of metadata we would need standards for metadata (like DDI for micro data on individuals).

Inconsistencies between aggregate and micro data outcomes:
This could be(come) a barrier for access to data. However, if -in general- these barriers are high and persisting the research community might look for alternatives. Currently the costs of collecting new data are very high in the Netherlands. But if these costs drop, e.g. because data collection costs decrease by using the internet, these researchers will start their own data collection.

NETHERLANDS [Stoop]:

Quality concerns may be a reason for not providing access to microdata. In practice, it may happen that quality concerns and not being able to invest time in assessing the quality or improving the quality might also be a reason for not providing access. Quality concerns might also be an easy excuse for not providing access. This should be avoided.

UNECE [BRUENGGER]:

NSOs can guarantee the quality of the microdata according to their own quality frameworks. However, the responsibilities for any results compiled by a university researcher on the basis of such microdata (unless this is done on the basis of a mandate by the NSO given to a researcher) lies entirely with the researcher; such results are therefore not part of official statistics. In my view, it is important to underline this very clearly in the paper, because sometimes researchers compile aggregates according to their own definitions that are close, but not identical to, the definitions, methods of aggregation and classifications used by the NSO in calculating results of official statistics. When research results are published, it is very important that the public is not mislead in believing that these are results that compete with those officially released, but are special constructs for a specific research purpose.

**[COST]**

*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.)*

NETHERLANDS [den Butter]:

The trade-off between the degree of privacy protection and the costs of using the data should, in a way, be the subject of a cost/benefit analysis. Changing preferences with respect to privacy protection could imply easing the guidelines for use so that its costs are lower. However, the risk of a loss of reputation and the consequent costs of for collecting data in future should always be part of that cost/benefit analysis.

NETHERLANDS [Dekker]:

Costs : In my personal opinion the research community (e.g. research council) should be willing to pay for these marginal costs.

FINLAND:

However, the division of costs between researchers and statistical agencies in improving the research use of micro data could be discussed more in the paper.

**[NSO SUPPORT TO RESEARCHERS]**

*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.*

*75. (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.*

NORWAY:

NSO should by any use of microdata be named as the origin of the data.

NETHERLANDS [Dekker]:

Supporting research community: I think this could be a win-win situation and it could increase the use of the (micro)data.

FINLAND:

Co-operation with the statistical staff is very important. Firstly, knowledge and advice on how to use and interpret the data are very essential for the researchers. Secondly, the feedback from the researchers helps in improving the applicability and quality of statistical surveys. However, this feedback should be gathered in a more systematic way.

GERMANY:

Paragraph 75. (d) It is very important to disseminate the results of microdata researches. The complexity of economic and social change and the progress made in science and information technology have led to a fundamental change in modern societies' need for data. Standardised tables published by the NSO may not be enough anymore for a sufficient description of modern societies. And also more disseminated scientific results enable to verify the objectivity and reliability of themselves and of already published results and therefore the validity of such data is improved. The German FSO has already taken a great leap in that direction. With the establishment of research data centres the collaboration with science has improved a lot. In the last years several results of research have been published, which are based on official microdata. The Federal Statistical Office of Germany even put out a series of books connecting official microdata and empirical research this year. In that series, researchers are able to publish dissertations or other important research work, which was produced by analysing microdata of the official statistics.

NORWAY:

75. (d) We consider this to be important and request that research based on microdata files are put in the public domain.

UNECE [BRUENGGER]:

75. (d) This is an issue that pertains to the "culture and values" of the different research communities. NSOs should receive copies of the results of the research activities derived from the microdata, whether these results are in the public domain or not.

UNITED STATES:

75. (d) Putting all research based on microdata files in the public domain is a very interesting idea to explore further. Researchers may or may not like it, but it seems fair to discuss it.

NETHERLANDS [Stoop]:

75. (d) Good point. A condition for getting access might well be that the results should be made publicly available.

## V THE PERSPECTIVE OF THE RESEARCH COMMUNITY

**[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.*

NORWAY:

We are in the middle of a discussion ourselves about what is covered by the term research. Up to now we have only included universities and research institutes. Our Statistical Act allows access to microdata for statistical use in research and public planning. This means that we also might include government and NGO to our list. Business and international organisations (except Eurostat) is present not included to get access to out microdata.

UNITED STATES:

In paragraph 17, the word “only” is missing. The paragraph should read as follows: “17. The research community is not only those who belong to universities. It may include researchers in other research institutes, government, business, NGOs and international organizations for example.”

UNECE [BRUENGGER]:

Paragraphs 17 & 28:

In para. 17, the paper states that “the research community is not (only) those who belong to universities. It may include researchers in other research institutes, government, business, NGOs and international organisations e.g.”.

On the other hand, para. 28 speaks of the “culture and value system of the research community”, implying that the whole community as defined in para. 17 is bound by the same culture and value system.

It is certainly the intention of all researchers to behave according to this culture and value system. However, the ability to comply with these values may be different for the various types of researchers that are listed in para. 17 above. I generally agree with the proposals of the paper concerning researchers at universities for which the statement in para. 28 is certainly applicable (although epidemiologists may have a different culture in their desire to reidentify individuals as opposed to social scientists). I continue to have doubts, however, to treat in exactly the same way “researchers” from government ministries and international organisations. The business case lies between the two, but it is certainly true that researchers in private companies have a different culture (at least in social science) from those working at universities.

The main problem is with the possibility of conflicts of interest for researchers outside universities, especially with respect to the “exclusively statistical use”. In particular, I think of transition countries, where this concept of “exclusively statistical use”, which should be defined in the paper, is relatively new and still difficult to implement for statistical offices (e.g. in refusing transmission of individual data to government units for non-statistical use). This fight for a coherent stand on this issue would be weakened if the same government units can have legal access to microdata for research purposes without additional safeguards compared to researchers at universities.

I would therefore make a plea to reformulate para. 17. The paper, in its present form, should be about access to microdata for researchers at universities. This is especially the case for para. 20 and 28. The other categories of researchers may either be discussed at the end of the paper in a new chapter (with specific proposals), or left to future work in analogy to enterprise microdata. For researchers in government units, I continue to be of the strong opinion that they can receive microdata from official statistics only if they are part of a “statistical unit” to which the confidentiality rules of the statistical law are applicable. Researchers outside these units should ask the statistical units in their own department/ministry (or directly the statistical office) for any statistical work that requires access to microdata.

In order to make the issue on exclusively statistical use being a condition for receiving microdata, I propose to add at the end of principle 3“…. that excludes any non-statistical use.”

NETHERLANDS [Vollering]:

The discussion paper defines the research community in a generic sense (item 17). The Social Sciences Council, however, distinguishes at least two research communities: the first one is the scientific research community and the second one is the research community that is financed with money from the market. The scientific research community, on the other hand, is financed with collective money and endorses the "Gedragscode voor gebruik van persoonsgegevens in wetenschappelijk onderzoek" or likewise codes of conduct. Its research is performed in the context of a university or a scientific research organization. The Social Sciences Council advises the Task Force on Confidentiality and Microdata to ensure that microdata collected and preserved by National Statistical Offices is fully accessible for scientific researchers that fall under these codes of conduct.

NETHERLANDS [Stoop]:

The research community is not only those … (point 17)

**[BENEFITS OF MICRODATA ACCESS]**

*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.*

NETHERLANDS [Dekker]:

(i) For this NSO's have longitudinal data (time series, panels). In the future event-based data and/or data on subgroups in the population will become available.

(iv/v) (referring also to the win-win in 16) The research community could do 'part of the job' i.e. analyse data, construct new data (data merging/enhancement)

**[RESEARCHERS OWN DATA COLLECTION]**

*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.*

NETHERLANDS [Dekker]:

The comment on this made in 19 (that these data will be of inferior quality) is a misperception (esp. if professional data collecting agencies carry out the job) and underestimates the risk of researchers starting up their own data collection. However, on a national scale these new data collections will be inefficient. In any case it is important that the NSO's are 'bureaus of standards': other data producers should use NSO's definitions and coding schemes.

*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.*

UNITED STATES:

How can we be sure of what evidence there is of this? In fact, if identifying individual respondents furthers research because, for example, it allows matching with other data, researchers may be tempted to do so.

**[CONFIDENTIALITY PROTECTION METHODS]**

*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.*

KOREA [Individual Researcher]:

Comments on overall issues of microdata:

The ultimate objective of producing statistical data (microdata) lies in its use. Doubts about the usefulness of censuses seem to be raised because pratical use of statistical data is insufficient. So I suggest that the statistical providers (NSOs and other statistical producers) treat the provision of statistical data from the users' perspective.

As a user of statistical data I have a couple of points as follows:

a. The range of available data is restricted
b. It is difficult to know fully how to access it
c. A lot of time is needed to access it (access procedures are complicated)

Therefore, I would like to say from a user's viewpoint that use of microdata is not satisfactory.

Though users anticipate swift and convenient use of data, the providers have relatively conservative attitudes. They say that privacy should not be infringed and the low quality of microdata may be a secondary reason.

In any case privacy should not be contracted. When there are probabilities for the violation of privacy, the availability of statistical data should be restricted and finally use of the data could be impossible.

Users and providers are in a contradictory position to each other with regard to the data accessibility, however, when microdata cannot be used, it is as valueless as the jewels hidden in the ground.

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

KOREA [KSS[1]]:

There is no room to object to having privacy be top prioritized, especially for the NSOs.

Basically, the KSS assumes that the principles be drafted by bench-marking OECD members' systems.

In the context of Korea's circumstances, the NSO should be empowered as a risk manager by having the initiative in this matter. We suggest that the NSO establish a method to serve each separate group of users or researchers by dividing them into 3 or 4 sub-groups. For this, it is necessary that the NSO have the full responsibility and that a regulation be enacted that has no exceptions in case of violation.

NORWAY:

It is a general impression that the tension between Statistics Norway and the research community is not so much about what types of data that are available to researchers, but more over issues like documentation of data sources, pricing policies, service level to researchers etc.

[1] Korean Statistical Society (KSS)

UNITED STATES:

There are some good points, but there is no solution.

AUSTRIA:

The proposals made in points 68 (balancing the public goods of privacy and right to information), 69 (resolving tensions between NSO and researcher perspectives) and 70 (business data) cannot be supported as in our experience the individual business data (referred to in point 70) must be handled with particular care.

NETHERLANDS [Vollering]:

The discussion paper emphasizes that these responsibilities exist and that there is to some degree a tension between them. The paper furthermore seeks practical solutions for problems that arise from this tension. The Social Sciences Council, being aware of these responsibilities and the legislation regarding privacy protection in the Netherlands, has set up a code of conduct for academic researchers in Dutch universities and other scientific research organizations. This code of conduct (in Dutch: "Gedragscode voor gebruik van persoonsgegevens in wetenschappelijk onderzoek", April 2003) translates the Dutch privacy protection law to the audience of scientific research professionals involved in microdata handling. On the formal side, the boards of the universities and scientific research organizations are responsible for upholding the code of conduct. Of course, the government, non-profit organizations and scientific research organizations can - and must - take all sorts of measures, i.e. technical, administrative and legal measures to avoid risks from incidents. By definition, however, measures cannot be solid proof entirely. From this perspective it is wise to move from risk avoidance to risk management (item 23).

**[RISK MANAGEMENT]**

*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.*

UNITED STATES:

Paragraph 23 recommends that NSOs move from a Risk Avoidance to a Risk Management strategy. That may be contrary to the intent of some statutory language.

**[INCREASED IDENTIFICATION POSSIBILITIES]**

*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.)*

NETHERLANDS [Dekker]:

Identifiable persons: As new data will contain events the chance of identifying persons will drastically increase. Therefore I would plea for safe settings, rather than safe data.

ARMENIA:

I am fully sharing the concerns related to the risks of microdata identification and would like to mention that for a small country the risk on anonymised microdata is rather sharp.

**[PUBLIC ACCEPTANCE OF CURRENT PRACTICES]**

*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.*

UNITED STATES:

Paragraph 25 notes that microdata access has not been an area of public controversy, but also that it has not been a topic for public debate. The general tone seems to minimize the concerns NSOs should have, but it notes at the end an untoward incident could provoke a debate -- the notion that one incident that receives media attention could impact respondent cooperation is also noted in 12.

**[TRANSPARENCY]**

*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.*

NETHERLANDS [Kooiman] :

Mention *informed consent.*

**[RISK MANAGEMENT]**

*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.*

ARMENIA:

(ii) In transition countries, where the culture and ethics of public service have not been established, it not excluded that administrative bodies will give a tribute to temptation to get microdata from researchers, as well as especially to finance researchers to get microdata.

NETHERLANDS [Vollering]:

(vi) The Social Sciences Council fully agrees to put some of the onus of responsibility for protecting privacy of respondents to the research community (item 27.vi). It advises the Task Force on Confidentiality and Microdata to stimulate that research organizations in other countries set up likewise codes of conduct.

UNITED STATES:

Paragraph 27(iv), states that NSOs "be completely transparent about the specific uses of microdata to avoid suspicions of misuse." This is possible in a controlled environment but not for public use microdata where uses are not limited or monitored.

NETHERLANDS [Dekker]:

(i) and (vi) In NL we have a set of principles (Gedragscode) for researchers (on how to deal with individual micro data). This code will work if data producers (NSO) and research funding agencies (research council) would penalise misuse, e.g. by not allowing this researcher (or the faculty) access to any data and/or submitting research proposals for grants for a number of years.

(iv) UK Data Archive asks researchers to submit their research proposal for (each) data use. This system works well. In the NL we (at the Scientific Statistical Agency) also have ideas to ask for a research proposal (very brief, 1 page) when they use data.

This information clarifies actual use of (which) data and can be used as feedback to the research community (e.g. by publishing the link to the paper(s)).

NETHERLANDS [Stoop]:

Several issues that come up in this paper have been discussed at the 19th CEIES seminar "Innovative solutions in providing access to microdata held in Lisbon on 26 and 27 September 2002". Recommendations from this seminar have been adopted by CEIES (the advisory body of Eurostat, comprising produces and users of statistics) at its 12th plenary meeting in Brussels, 22 November 2002. For several reasons the follow-up of these recommendations has been delayed.

Central in these recommendation is how to solve the dilemma between risk avoidance, which can only be done by not making micro data available for research purposes, and making efficient use of public money and of respondent efforts by putting their data to good use in (policy) research outside official statistics.

The CEIES opinions on confidentiality might provide a further basis for and elaboration of the paper, as they both clearly seem to aim for making microdata available for research while keeping confidentiality risks at an acceptable level, a target of which I thoroughly approve.

See the following website for the minutes of the CEIES plenary meeting and the text of the recommendations: http://forum.europa.eu.int/Public/irc/dsis/ceies/library

Select Plenary meetings, then 12th Plenary Meeting, then English documents, and finally Minutes. See item 6, and Annex B.

GERMANY:

Moving from a risk avoidance to a risk management strategy seems to be just the right way to discover a good balance between researchers' access and confidentiality protection. The German Federal Statistical Office with it's research data centre approaches that aim in giving a part of disclosure risk back to the researcher by contract agreements additional to methods of data anonymisation and technical means of disclosure control, like remote access and access on safe scientific workstations in data labs. Sanctions for the breach of confidentiality can be the loss of scientific reputation, futurous banning of the researcher from microdata access or in very severe cases also mulcts or even imprisonments. A combination of technical and juridical

security issues with data anonymisation techniques makes it possible to achieve a good research environment by full protection of the respondent. If the surroundings of data access are safer, i.e. a safe scientific workstation or a binding contract, the change of the data itself for anonymisation purposes can be lower.

NETHERLANDS [den Butter]:

It is essential that Privacy Commissions underwrite the guidelines which, if needed, should be adapted or elaborated to meet national standards and preferences. In case of misuse or of other problems caused by analysts of the data, the NSO cannot be held responsible for organising the use of the data in an incorrect manner. In that case it is hopefully perceived as an incident, which will do less harm to the reputation of the NSO in guaranteeing confidentiality.

NETHERLANDS [Stoop]:

A combination of transparency, codes for researchers (possibly to be developed by the ISI), setting up penalties for violations (rather than prevent violations by precluding access) and remote facilities might do the job. The research community generally seems to be wary of distortion of the data, although this might be useful for test data.

KOREA [KSS]:

The basic principle should be that the NSOs have the initiative in all decisions regarding use of microdata files. However, to ensure NSOs move from a risk avoidance to a risk management, there should be discussion channels or forums for a long time for the users or researchers to express their critical opinions in this field.

KOREA [Individual Researcher]:

To define availability of data by classifying users:
It is to divide users into several groups and impose different restrictions of access according to the type of each group. It is a method that general statistics can be accessed freely on the web and the more detailed or privacy protection required data is accessed through respective restrictions. For example, the range of access can be broadened to the official statistics providers and researchers of a specified field. However, it can be assumed that the classification of users into several groups may not be easy.

To define availability by the objective of the use of microdata:
The users of microdata should be defined into two groups. One group is to use microdata as a frame of producing new statistics and the other is to use microdata directly in the various areas, for example, using data collected from the Household Income and Expenditure Survey for the economic and social analysis.
And when government agencies access data for public goods or data is used for the production of government approved statistics, the access right of the agencies should be prioritized or broadened.

In the cases above, a written oath is needed so as to impose full responsibility to the users.

Research to protect privacy:
It is necessary to continue to develop techniques to protect data privacy while permitting use of data at one's convenience like data perturbation.

I would like to suggest as follows for NSOs to provide convenient access of microdata for the researchers while maintaining confidentiality protection.

**[RESEARCHER RESPONSIBILITY]**

*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.*

UNITED STATES:

Paragraph 29 discusses techniques to put the risk back on researchers, most of which are already in use. An interesting idea in paragraph 29(i) is having researchers prove their bona fides and demonstrate the public benefits of their research. While good in principle, this would seem to be difficult to operationalize and apply to the wide variety of content areas that are covered by statistical agencies.

In paragraph 29, we would suggest spliting subparagraph (iv) into (iv) and (v) by moving "Undertaking legal action." to (v). It is one of the most important ways of reallocating the risk to the researcher and should stand-alone.

**[COMBINATION OF MEASURES]**

*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.*

NETHERLANDS [Vollering]:

The Dutch Social Sciences Council agrees that the research community should not have automatic right of access (item 30).

**[CONTINGENCY PLANNING]**

*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.*

UNITED STATES:

Paragraph 31 discusses contingency planning in the event that microdata access becomes an issue for debate. The paper offers good ideas, but does not acknowledge as it did earlier that damage to the system from a negative incident could likely not be repaired.

## VII ANONYMISED MICRODATA FILES

**[PUBLIC USE FILES]**

*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.)*

*75. 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.)*

GERMANY:

75. (a) The question is, if users of Public Use Files should also sign a contract before using the file. That would have the advantage of exactly knowing who is using the data (that could be important especially in copyright issues), but on the other hand the file is by definition absolutely anonymous and so a contract should not be necessary. In Germany the necessity of a contract for Public Use Files is not seen. A standardised Public Use File, i.e. the 25%-subsample of the statistics of public assistance can be purchased as a standard product of the German Official Statistics like the Statistical Yearbook. A Public Use File for statistical training purposes like the Campus File is even freely downloadable from the Internet.

KOREA [KNSO[2]]:

75. (a) The provision of Public Use Files should be discouraged unless there is some form of undertaking by the person accessing the file, as in the case of the confidentiality risk.

NORWAY:

75. (a) Our legislation do not admit Public Use Files. Access to microdata is restricted to researchers and public planners.

UNECE [BRUENGGER]:

75. (a) I would not discourage public use files in general (for household and individual data). It may be wise, however, to release such files only after a certain period of time, so as to decrease the possible accuracy of the information at the individual level for any person who might be tempted to reidentify persons for other than statistical use.

UNITED STATES:

The paper identifies issues where the authors had disagreements.

[2] Korean National Statistical Office (KNSO)

75. (a) suggests that the provision of Public Use Files be discouraged unless there is some positive action on the part of users of public use microdata to assure they will not violate confidentiality. The Census Bureau does not require any type of positive action (contract or agreement) but other U.S. agencies do. This should be left to each NSO and not be a recommendation of this paper. Each NSO understands the dangers in producing such files and each is in a unique position to make a decision about the availability of such files. Recommending that an appropriate amount of research be made to ensure the anonymity of the microdata is appropriate.

*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.*

EUROSTAT:

The use of Public data file could be a possibility to decrease the administrative and organisational burden on NSOs.

NETHERLANDS [Dekker]:

PUF: I would distinguish between research and training. The latter doesn't need real data (i.e. data can be permutated (PRAM), or disturbances can be added)

## VIII REMOTE ACCESS FACILITIES

EUROSTAT:

Remote access facilities seem to provide a good solution and will be considered by Eurostat for the future. It seems however there is still some work and studies to be done to reach the goal. There is a need for further methodological work on the disclosure risk of output of statistical analysis and technical work on how to implement automatic checks.

FINLAND:

Providing protected remote access, as suggested in the paper, could be an alternative to on-site access. Remote access facilities are likely to have an increasing role in the use of confidential micro data. After the set-up, remote access is cheaper to maintain and easier to control with appropriate technical solutions. It also increases the equality of researchers both in geographical and funding terms. However, the question arises whether the high set-up costs should be covered by public funds.

GERMANY:

Remote access is the most comfortable way of microdata access for researchers and also for statistical offices. It seems to be the future of microdata access. Remote access is international (the researcher and also the employee of the statistical offices can stay at their workplace), it is time-saving (especially for automatic remote access solutions there is almost no delay between processing of data and result) and at least it is safe (every process in the working chain can be controlled). The greatest disadvantages are the high prime cost especially for an automatic client-server solution and the high amount of work and time for the non-automatic solution.

The Federal Statistical Office of Germany is currently considering two complementary ways of remote access. On the one hand, we are testing a terminal server solution similar to the danish system in which remote users could analyse project-specific Scientific Use Files by means of standard analysis software like SPSS or SAS. Before puting this in practice, some legal aspects have to be discussed in the next time.On the other hand, we are developing methods of confidentiality protection for the results of statistical analyses of microdata which

could be applied in the context of a system for remote access in which the user can analyse but not see the original, that is, not anonymised data (see also the attached working papers). These ideas will be discussed with prospective users and we are interested in the experience other NSOs in that field. Part of the papers will also be presented to the COMPSTAT symposium in Prague this August. Where the first approach gives users more comfort, the second approach gives them more precision especially in the case of business or regional data.

KOREA [KNSO]:

Agrees to it and security programs should be set fully against hackings which may happen with the development of information technology.

KOREA [KSS]:

Agrees to it.

NORWAY:

We agree with the paper that RAF are important and useful, and that the usability if these arrangements needs to be tested. We want to make steps to use RAF as a supplement to ordinary licensed files.

SLOVENIA:

We had a very fruitful discussion about remote access facilities. We have to say that the medium-term perspective in Slovenia is to establish a facility similar to the Danish one. This idea was presented and researchers supported the idea strongly with slight concern about the possibility of the introduction of some special software they use for specific analyses. The statistical office strongly supports the idea of the remote access facility, especially because of the transparency of dealing with micro data. We see the transparency of dealing with micro data as a key element for public support to dissemination of micro data. Researchers strongly support the idea of linking data sets, which is the reality in Slovenia and done exclusively by the statistical office.

UNITED STATES:

Yes, it seems that Remote Access Facilities (as defined in this paper) will play an increasingly important role.

NETHERLANDS [Stoop]:

Remote access and remote execution can be useful means of providing access to micro-data. The facilities should be user-friendly.

AUSTRIA:

Remote access facilities proposed in point 73 might be a possibility, although the input with regard to bureaucracy, control and monitoring will be enormous, and implementation possibilities had to be tested beforehand with particular respect to the cost benefit effect. In any case, provisions would have to be taken to keep data confidential, and the prior consent of NSIs will be essential in the question of granting access to microdata for scientific purposes.

*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.*

UNITED STATES:

Paragraph 37 states that only Denmark uses Remote Access Facilities where the researcher performs analysis and can immediately see the answer. The U.S. Census Bureau's AFF Tier 3 Query System also provides this capability.

*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.*

NETHERLANDS [Dekker]:

Sharing experiences: This would be very useful for NL, since we are setting up a new data services organisation. In this respect NORFACE could also be mentioned: an EU initiative of (currently 7) social sciences research councils to cooperate on EU level (another setting would be the European Research Council).

## IX DATA LABORATORIES

No comment

## X ENGAGING AS TEMPORARY NSO STAFF MEMBER

No comment

## XI BUSINESS DATA

FINLAND:

The confidentiality of business data should be guaranteed. However, setting up and maintaining functional on-site research facilities for business data is quite costly for the NSOs and the researchers. In addition, much information is publicly easily available from other data sources, which decreases the incentives for data identification. As a consequence, the costs of protecting the data may exceed the benefits of disclosing the data.

**[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.*

GERMANY:

Business data is to be treated different than personal or household data. Due to the fact, that the number of respondents is much lower in business datasets and their characteristics are more identifiable, the anonymisation level of this microdata has to be higher. The Federal Statistical Office of Germany has experts who are working together with experts from the empirical sciences and representatives of the statistical offices of the Laender in a project titled „De facto anonymisation of business data“. In this project several methods of data protection are tested on business data with the aim to put out business microdata files for scientific access. In that context also data perturbing tools like microaggregation are used. Perturbing methods are never used for person or household data in Germany. But nevertheless, all the other principles of data access as stated above, are valid as well. You can also manage to give some of the disclosure risk back to the researcher by contracting. Of course it is also possible to use technical security tools, like remote access or data laboratories to protect confidentiality. Due to the fact, that the respondents are more easily identifiable, one could think about some kind of waiver-concept for confidentiality protection of businesses. That means: The respondents (resp.: Companies) could be asked for permission to provide

their data for scientific research. For a proper risk management strategy that would mean, to not only put some risk back to the researcher but also put some risk back to the respondent.

POLAND:

Referring to the item 70 of the elaboration. As regards the item 11 of "Business Data" I maintain my previous comments since, I am afraid, the CSO cannot agree on dissemination of micro-data on economic activity of business entities, including agricultural holdings, due to a big difficulty connected with data "anonimisation". Simultaneously, a too far reached anonimisation will considerably limit the cognitive value of such data. It seems to me that it would be difficult to reach in this matter an agreement on the part of small entities, and even single protests willingly made public in mass media can bring serious damage and can increase the rate of refusals of respondents to fulfil questionnaires. Therefore, we should not support the item.

SLOVENIA:

About business and personal data, we do not see particular difference in the sensibility of data. From the point of view of the small market, in Slovenia it is particularly demanding to protect business data.

UNITED STATES:

Yes, access must be different for business data.

*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.*

KOREA [KNSO& KSS]:

Paragraphs 46 & 47: Agrees to it.

UNITED STATES:

Treatment of Business Data: Paragraphs 46 and 47 discuss business data. In addition to the concerns of confidentiality directly affecting an information provider, we would note that a country may have economic competitiveness (and possibly even security) issues with respect to sharing identifiable business data with researchers in other countries. A country may have an interest in ensuring that certain information is not made readily available to foreign interests. For example, General Motors has an interest in ensuring that information it reports to the U.S. Government is not made available in identifiable form to its competitors (domestic or foreign), but the U.S. Government may also have an interest in ensuring that

foreign vehicle manufacturers and their governments do not have access to identifiable information that could be used in planning economic or competitive strategies.

FINLAND:

As regards Statistics Finland, the research laboratory located at the Business Structures unit has had good experiences about the research use of business data. The researchers have been responsible and reliable, and data confidentiality violations have not been discovered. In several projects the data have to be tailored for the purposes of the researcher.

EUROSTAT:

It is suggested that access arrangement might differ between business and individual data but the principle should remain the same. However, it must be recognised that the sensitiveness and the risk are intrinsically different. The paper does not explore the "deconfidentialisation" of publicly available data, especially for large enterprises.

SLOVENIA:

About business and personal data, we do not see particular difference in the sensibility of data. From the point of view of the small market, in Slovenia it is particularly demanding to protect business data.

NETHERLANDS [Kooiman]:

Paragraph 47: first bullet: less --> more

AUSTRIA:

The proposals made in points 68, 69 and 70 cannot be supported as in our experience the individual business data referred to in point 70 must be handled with particular care.

NETHERLANDS [Stoop]:

Data Laboratories might be a good solution for providing access to business data.

## XII SPECIAL ISSUES

**[LINKED DATA SETS]**

*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.*

NETHERLANDS [Dekker]:

linked data sets: A new option/development is linking individual data and geographical data. This combination largely increases identification.

**[EQUAL ACCESS]**

*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.*

*75. (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?*

NETHERLANDS [Dekker]:

Equality of access: Everyone is allowed to ride a bike; for a car you need to have a license (revealed expertise) and for a large truck, boat or airplane the training requirements are even higher.

Statistical (micro) data also come in several degrees of complexity: graphs, tables, records, panels, events, with increasing need for training/expertise.

Hence it would make sense to 'require some expertise' before giving access to complex data.

GERMANY:

75. (b) Equality of access to information about the society is one of the highest principles in modern societies. Following that, also access to data describing social processes should be equal. But on the other hand there is also a right for data protection of the individual, what makes it indispensable, to have some means of control on the people working with official microdata to protect the individual.

KOREA [KNSO]:

75. (b) Equal access should not be applied to all data. Accessible data should be designated in advance according to type and users have equal access within the designation.

The discretion of NSOs should be confined to the provision of regulations according to the types of data.

UNECE [BRUENGGER]:

75. (b) Equality of access is a principle that is only applicable to results of official statistics, as well as to metadata about official statistics (activities and results). With the exception of public use files, it would be dangerous to extend this principle to protected microdata. Access to microdata has to remain at the discretion of NSO, based on transparent criteria and fair procedures.

NORWAY:

75. (b) It is a general view that the principles of the NSOs provision of microdata should be transparent, and any discretion in principle should be formalized into transparent guidelines.

UNITED STATES:

75. (b) NSOs need discretion on who should have access to their data.

NETHERLANDS [Stoop]:

75. (b) Discretion to NSOs? As mentioned before, this might lead to subjectivity or to setting a low priority to providing data access due to time of money constraints.

AUSTRIA:

75. (b) Cannot be supported.

KOREA [Individual Researcher]:

Though equal access principle may not be adequate with the use of microdata, since all users do not try to access it for the public good, an appropriate level of equal access should be ensured as long as data confidentiality is secured.

*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.*

UNITED STATES:

If an organization wants the capability to release identifiable information to researchers, the information provider should consent to that condition at the time of collection. (Paragraph 50 indicates that informed consent is not feasible for large statistical collections – we would disagree.) We would argue that passive informed consent is possible where respondents are informed at the time of collection of the possibility of sharing data with researchers approved by the organization collecting the information. If an organization is unwilling to provide such notice as a condition for requesting information (e.g., concern that possible sharing of information may cause unease with respondents and negatively affect response rates), then the organization is, in effect, admitting that information providers may not be comfortable with such sharing arrangements.

## XIV PRINCIPLES

KOREA [KNSO]:

The discretion of NSOs should be confined to the provision of the regulations of aims, methods and procedures of the use of data according to the types of data. The discretion should not be for any authority to decide whether to permit access to an individual data file that is applied to for use.

The KNSO comments that the draft principle be made more concrete and comprehensive.

KOREA [KSS]:

Agrees to it.

*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."*

POLAND:

The comment notified in my letter of January 2004 in which I paid attention to the fact that proposed solutions concerning the dissemination, by statistical services, of data for scientific purposes are not consistent with item 6 of "Fundamental Principles", is still valid. The principle states explicitly that "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"**. You also stresses that "Any principles for microdata access must be consistent with this Fundamental Principle". However, proposed solutions seem to be inconsistent with the statement. It would be advisable to explain whether we accept the fact that statistical analyses used to support scientific surveys constitute a logical part of statistical uses (which seems to be a little far-fetched), or, taking into account the importance of using the statistics for scientific surveys, the statisticians will tend to modify this statement of the item 6 of "Fundamental Principles". None of the "principles" suggested in the item point gives an unambiguous answer to this doubt. I should also point out that article 10 of the Polish "Law on Official Statistics" does not give a free choice in this matter either, unless we consider scientific uses as equal as statistical analyses. I do not know who could be entitled to formulate such a comment, as well as, how and by whom the comment could be overused.

NETHERLANDS [Dekker]:

Statistical purposes: Does 'statistical purposes' include scientific purposes? (Not only relevant for the science community but also for the NSO's if they want to go beyond statistics.)

KOREA [KNSO]:

The aims of use of microdata should be classified according to the types of data.

**[PRINCIPLES]**

*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.*

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

FINLAND

It is a good objective to create some common principles and guidelines for improving the use of micro data, as part of the "Risk Management strategy" mentioned in the paper, to reduce the tensions between the NSO and researcher perspectives. However, the concepts in the draft principles should be defined carefully, concerning for example the definition of research. Each country should have the right to assess each project separately and to use the ethics committee to evaluate unclear cases. It is also important to create clear rules and penalties of which the researchers are informed during initial training.

SLOVENIA:

I would like to inform you that the Statistical Office of the Republic of Slovenia strongly supports the idea of producing a set of principles and associated guidelines for managing access to micro data by the research community at the international level.

In the paper there are some issues the statistical office supports but the research community does not have a clear opinion about it. For example, the use of micro data files should be banned for non-statistical purposes, discretion of access to micro data should be based on the need to know principle, etc.

UNITED STATES:

In paragraph 53, we would suggest changing "each" to "they," i.e., "They are discussed in the following paragraphs."

EUROSTAT:

Principles 2 & 3:

The use of micro data for non statistical purpose should be banned given the statistical purpose is the main reason for the existence of a statistical confidentiality regime. The statistical confidentiality regime appears to be rather favourable for developing official statistical activities in comparison with the standard framework for the respect of the right for privacy. This specific status should be maintained

In addition, we would like to stress the importance to have common agreement on disclosure risk measure of micro data taking into account the new technological development. Ideally this measure should have the support of Privacy Official and Member States'. This measure

could help to identify relatively low risk micro data file which could be release much easier (public use file).

ARMENIA:

I think that it would be desirable to consider as a subject of discussion the removal of proof power of microdata received by the NSOs based on the law, e.g. that microdata could not be a basis of administrative decisions and used as a proof in the court.

GERMANY:

Paragraph 75. (c) The data required to analyse and shape modern societies must in particular provide information on social sub-groups and allow to perform analyses of economic and social change on the basis of longitudinal data. Due to the changed information demand, it is no longer sufficient to publish results only in the form of tables. To meet the requirements in terms of methodology and content, it is necessary to present statistical data also in a way meeting the data demand of the scientific community. This includes providing access to anonymised and non-anonymised microdata which allow to perform more varied analyses.

KOREA [KNSO]:

75. (c) It can be banned according to the types of data. Public Use File can be used for non-statistical purposes.

NORWAY:

75. (c) Our legislation only permits use of microdata for statistical purposes (in research and public planning)

SLOVENIA:

75. (c) In the paper there are some issues the statistical office supports but the research community does not have a clear opinion about it. For example, the use of micro data files should be banned for non-statistical purposes, discretion of access to micro data should be based on the need to know principle, etc.

UNITED STATES:

75. (c) There should be no non-statistical uses of the data.

NETHERLANDS [Stoop]:

75. (c) Non-statistical purposes? What are non-statistical purposes? Data should not be used for identifying individuals and disclosing information or acting on specific individuals. On the other hand, a commercial organization might do statistical research for its own purposes. Is that a non-statistical purpose?

AUSTRIA:

75. (c) In principle, statistical data should be used for statistical purposes, only, except in case that another purpose is determined by legal norms.

UNECE [BRUENGGER]:

The main problem is with the possibility of conflicts of interest for researchers outside universities, especially with respect to the "exclusively statistical use". In particular, I think of transition countries, where this concept of "exclusively statistical use", which should be defined in the paper, is relatively new and still difficult to implement for statistical offices (e.g. in refusing transmission of individual data to government units for non-statistical use). This fight for a coherent stand on this issue would be weakened if the same government units can have legal access to microdata for research purposes without additional safeguards compared to researchers at universities.

In order to make the issue on exclusively statistical use being a condition for receiving microdata, I propose to add at the end of principle 3"…. that excludes any non-statistical use."

75. (c) This question should not have been asked. The fundamental principle is clear: exclusively for statistical use. If the majority of task force members are of a different opinion, we should be consequent and ask for a reformulation of this principle. I am strongly against opening this door, because this would undermine the distinction between statistical and non-statistical use, which is one of the main arguments for institutional specificity of official statistics within governments.

Principle 4:
To my knowledge, all contractual arrangements with researchers that underpin access to microdata from official statistics include a clause prohibiting the redissemination of microdata to "third parties". The reason is that the statistical producer who owns the data (and carries the responsibility for the respect of the confidentiality rules for these data) has to know, at any time, who has a legitimate access to confidential microdata. This overview is impossible if researchers can get access by asking other researchers, rather than the statistical office, for access. Therefore, I propose that this ban on redissemination be made explicit as part of principle 4.
Researchers working together on a project should be given access as a group (or as an organisational unit). This is another reason why the different types of researchers have to be distinguished.
Restrictions on redissemination are of course not applicable to public use files.

UNITED STATES:

Informing Information Providers of Possible Research Uses:
The principles (paragraph 53) do not fully address an issue that requires more emphasis, i.e., the ownership of the information.

NETHERLANDS [Dekker]:

I would agree with these principles.

GERMANY:

**Principle 1:** The Federal Statistical Office agrees to that principle. Access to data from the official statistics for research purposes should be out of the question today.

**Principle 2:** In Germany data access is regulated by the Federal Statistics Law: The German Federal Statistics Law in Article 16 (par. 6) BstatG for example legalises project-related research access for researchers from independent research institutions. They are given a so-called privilege of science and have access to less anonymised microdata. What might be necessary next is an international regulation going the same way. But not only data access should be regulated, above all ways of internationally prosecuting data snoopers should be found, so that an additional protection of the individuals rights by contract-binding of international researcher is possible. That would make it much easier to treat an abroad researcher equal to a German researcher in questions statistical disclosure control.

**Principle 3**: That principle is already reality in Germany, but as you can see above, there is still a difference between national and international research.

**Principle 4:** For principle 4 one has to think about the data protection of the researcher. Is it really a good idea to publish researchers' names and maybe also some other facts about this researcher on a web site? It should be enough, to let data producers check the bona fide of a researcher before giving him or her access to official microdata. But if publishing researchers on a web site really helps to protect confidentiality, the researchers should at least be asked, if they agree to the publishing of their private data.

NORWAY:

It should be possible to simplify the requirements by claiming that the result of the research are to be present as statistics, and that publication comply with the rules for official statistics. The views expressed in the report correspond well to the suggested principles.

UNITED STATES:

The draft principles in Section 13 should generally be supported. Some may have concerns about taking Principle 4 as far as proposed.

Principle 4:
If a person/business/etc. provides information to an organization through a survey or an administrative filing, the person/business/etc. has an ownership right in the identifiable information, how confidentiality is protected, and how the identifiable information is used. For example, if a person reports income to a statistical agency in a survey or to the Internal Revenue Service (IRS) as part of tax filing, that person has an interest in that information.

?? If the organization tells the person that it will protect confidentiality, the person should be able to expect that organization to take appropriate actions to do so, consistent with any pledge made. (The paper does a good job of addressing this issue.)

?? Unless the organization explicitly tells the person that the information will be shared with nonaffiliated researchers and used for other purposes, the person should be able to assume that identifiable information will not be shared. The person will have certain expectations about how the organization collecting the information will use it, and decisions on whether to participate are influenced by that knowledge. Unless informed otherwise, the person would view it as a violation if the organization later decides to share identifiable information with researchers for what the organization unilaterally decides is the "public good." If the organization informs the person at the time of collection that the information may be provided to independent researchers for projects that are deemed to be in the public interest, then the person can consider that in deciding whether to participate.

For example, a person can choose either to provide or not to provide income information to specific organizations such as a statistical agency and the IRS (and may choose to tell the IRS and not the statistical agency). However, providing the information does not confer a right upon either organization to then provide identifiable information to researchers when that organization decides it is in the public good, regardless of what confidentiality promises the collecting organization has made.

NETHERLANDS [den Butter]

It seems warranted to come eventually to a uniform world wide set of guidelines for the use of statistical data of NSO's by individual analysts. This could be done under the auspices of the UN, like in the case of the standards for the National Accounts. The paper provides a first set-up for such guidelines. The advantages for such uniform guidelines are that (i) there are economics of scale for both the producers and the users of the data: the wheel does not have to be invented again for each country and users will have less transaction costs, especially when making international comparisons by analysing databases from various NSO's; (ii) such common well-designed guidelines will be easily accepted and underwritten by national Privacy Commissions.

NETHERLANDS [Stoop]:

The draft principles seem sound. One concern, however, is that it will be easy to "hide" behind quality concerns or the inappropriateness of providing access to specific persons or

institutions. The discretion of the NSOs might unnecessarily preclude access, or postpone access until data are no more relevant to, for instance, policy purposes.

AUSTRIA:
We basically support principles 1, 3 and 4 of item 71 but not principle 2.

**[PRINCIPLE 1]**
*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.*

**[PRINCIPLE 2]**
*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.*

UNITED STATES:
Paragraph 55 says that arrangements made under Principle 2 should be cleared through the NGO acting as watchdog. Countries have differing rules about the participation of such NGOs. Although Principle 2 is an extremely important principle, the role of the NGO should be left to the individual countries.

**[PRINCIPLE 3]**
*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.*

UNECE [BRUENGGER]:
Second sentence: Any request for microdata subject to statistical confidentiality for non-statistical use, even by a court or a competition surveillance authority based on their own laws, is incompatible with the fundamental principles and therefore also with most statistical laws. I would therefore propose to drop the reference to “legal” requests for such use. This should be treated as the collision of two laws, and directors of NSOs should be encouraged to have such an issue brought up to the highest level of jurisdiction before granting access.

UNITED STATES:
Paragraph 56 says that it may make sense to apply a *compatibility test* for Principle 3. Although we agree with Principle 3, not everyone may agree with a requirement for a compatibility test. We assume such a test means that the research must support the mission of the NSO originally collecting the data. Some of the most useful research will be accomplished as a result of using data collected for one purpose for a completely different purpose. In fact, this may sometimes be the strongest argument for microdata use being in the public’s interest. In the case where a country has a decentralized statistical system, (addressed in paragraph 66), the paper states that the four principles should apply. This is probably generally true, but the examples of application of the principles in paragraphs 54 through 57 do not apply for these purposes. For example, it may be that microdata collected by one statistical agency can be used by another statistical agency for sampling frame purposes or other purposes that provide significant efficiency gains and corresponding budget savings that are in the public interest. These kinds of activities are contemplated in paragraph 66, without recognizing that they may not pass the compatibility test.

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

KOREA [KNSO]:

The Public Use File should be accessible to everyone regardless of the aim of use.

## XIV DATA LINKING

EUROSTAT:

The principles regarding data linking are rather general. The identification risk through statistical matching should be further worked out. Elements for its assessment should discussed in the paper as well as the actual risk when data are release to researchers

GERMANY:

The linking of data sets is treated very restrictive in Germany. By jurisdiction of the Federal Statistics Law (§13 BstatG 1987) it is not allowed to match different datasets of the official statistics. Even employees of the official statistics are only by way of an exception allowed to link different data. But to increase research quality there are efforts made to change the Federal Statistics Law and allow the linking of certain data for research purposes.

KOREA [KNSO]:

The proposed principles for linked data sets are appropriate.

For the purpose of confidentiality protection the authorization of rights to access to microdata should be rigidly restricted.

UNITED STATES:

The issue of transparency for linked data sets may be more sensitive and raise concerns about public perceptions.

In paragraph 57, the author indicates that "The NSO web site is an effective way of complying with Principle 4. This requirement would not apply to Public Use Files."

NETHERLANDS [Stoop]:

Linking data sets will contribute to the quality of research This should be done within the framework of the four principles.

AUSTRIA:

The linking of data sets should be a prerogative of NSIs, which are subject to particularly strict confidentiality and security conditions. For this reason, the linked data sets produced by NSIs should not be passed on to external third parties.

*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.*

UNITED STATES:

Paragraph 59 states that the statistical agency has to be the custodian of linked files. This may not be desirable or always legally possible. For example, a public use file might be linked to a users' proprietary microdata file. The result may provide important research results that could not have been achieved if the linked file had to be provided to the NSO.

*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.*

UNITED STATES:

Paragraph 60 underplays the risks of sharing linked data files with the public because it does not discuss the possibility that the source agency for the second (linked) file, if not the NSO, might find it easy to reidentify respondents to the NSO's survey

## XV ACCESS BY INTERNATIONAL RESEARCHERS AND INTERNATIONAL AGENCIES

EUROSTAT:

Regarding access by non national researchers, the current EU legal framework foresees that non EU research bodies must be declared admissible to get access. However their requests are put on an equal footing with other EU non research organisations.

GERMANY:

What might be necessary next is an international regulation going the same way. But not only data access should be regulated, above all ways of internationally prosecuting data snoopers should be found, so that an additional protection of the individuals rights by contract-binding of international researcher is possible. That would make it much easier to treat an abroad researcher equal to a German researcher in questions statistical disclosure control.

UNITED STATES:

Section 15 on access by international researchers and international agencies raises good issues.

*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.*

FINLAND:

(iv) Allowing micro data access for international researchers could improve the quality and comparability of the research conducted. So-called demo data sets (or dummy files) could be distributed to increase equal opportunities to use the data. International projects and co-operation on sharing experiences and best practices in using micro data are valuable.

NETHERLANDS [Kooiman]:

OK, but the amount of information that could be used by researchers (eitehr deliberately or accidentally) for a reïdentification is much less abroad than in the home country. To my opinion this effect dominates the other, so that, generally speaking, the disclosure risk is much less when sending microdata to researchers abroad. (Much less population knowledge)

NETHERLANDS [Dekker]:

International access: National (or European) Research Councils could obstruct national researchers if they would misuse data (e.g. no access to grant procedures or national data). The key point is that if data access goes through one (national) office (cf. one-stop-shop) the power of this office increases, since any misuse has implications for the researcher (misuser) w.r.t. access to all data sets.

KOREA [KNSO]:

Agrees to it and a national system needs to be established to discipline foreign offenders.

KOREA [KSS]:

Agrees to it and international researchers should be grouped as national users. The International law system in this field should be established to treat matters that may arise.

NORWAY:

We have recently clarified that our Statistical Act covers this solution. We are able to release data (both confidential data and AMF) through the NSO of the country of the researcher, provided that this country has adequate legislation to protect the confidentiality of the microdata. This includes use of RAF to provide AMF to international researchers.

NETHERLANDS [Stoop]

Considering the importance of cross-national comparability, the increase of international research projects and the increasingly open structure of the EU, it seems advisable to consider not to differentiate between national and international access. One additional point is that – whereas action against infringes in other countries might be more difficult – disclosure might also be more difficult in other countries.

AUSTRIA:

Concerning international access, we support principles 1, 3, and 4, but not 2.

UNITED STATES:

Penalties for Foreign Researchers:

The discussion of sharing information with researchers in other countries raises practical and legal concerns of how to pursue enforcement actions against foreign researchers who violate data sharing arrangements. While an organization relies quite heavily on a researcher's ethics when sharing data, the potential penalty for violations is an important component. It would seem to be very difficult to enforce legal sanctions against a researcher in another country who violates confidentiality principles or uses data inappropriately (see paragraph 62). As the paper admits more than once (e.g., in paragraphs 12 and 25), any violation of confidentiality can have serious repercussions for statistical organizations. (Paragraph 28 acknowledges that researchers have disregarded the controls inherent in microdata access arrangements.) An organization needs readily available penalties as a tool to help ensure compliance with any sharing agreements. Sharing with researchers in other countries raises concerns about enforcement and this issue should be more fully addressed. Also not addressed are questions of how the shared data are to be safeguarded in the computers of the foreign researchers, and what is to happen to the data in these computers once the project is completed.

*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.*

UNITED STATES:

In paragraph 63, is a distinction being made between public use files in (i) and licensed files in (iii)? We assume that paragraph 64 from the second sentence on applies to licensed files. If so, we agree with the statement.

*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.*

GERMANY:

The thoughts on access by international researchers outlined in Section 15 are already possible in Germany by now. International researchers can purchase Public Use Files and can also have access by controlled remote data processing or safe scientific workstations . If these principles were obtained in all member countries that could be a very good beginning. In a long run a researcher should be treated as a researcher no matter where he or she comes from.

UNITED STATES:

Restrictions on Access for Foreign Researchers:
We would note that U.S. agencies do not currently have a consistent process for making data available to external researchers. For example, in the case of the Energy Information Administration (EIA) in the U.S. Department of Energy, EIA has a Fellowship program with the American Statistical Association that can be used to provide access to microdata. Under this program, the researcher is asked to sign agreements and becomes a "contractor." Because of security concerns, including data security, the Department of Energy is changing its policy to require that all contractors be U.S. citizens. This new policy means that EIA will not be able to use the "contractor" approach to allow a foreign researcher access to its data.

*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.*

NETHERLANDS [Dekker]:

Remote access: Fully agree with this point.

## XVI SHARING OF MICRODATA BETWEEN STATISTICAL AGENCIES

No comment

## XVII ISSUES FOR DISCUSSION

**[TWO PUBLIC GOODS]**

*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.*

FINLAND:

Balancing of the two public goods, confidentiality protection and researcher access, is a very central issue. It is important to improve possibilities to use micro data for research purposes but without causing any risks concerning statistical confidentiality. In this light e.g. remote access with effective data security arrangements may be a good solution. However, the division of costs between researchers and statistical agencies in improving the research use of micro data could be discussed more in the paper.

NETHERLANDS [Stoop]:

In a way respondent efforts could also be considered a public good, acknowledging that these efforts deserve optimal use of the data provided by the respondent, rather than locking them up.

GERMANY:

Paragraphs 67& 68:

The German approach to a better informational infrastructure with implementing research data centres and the legal framework giving access to German official microdata is funded on the principles of these two public goods: Freedom of science with research access to official microdata and freedom of the individual with confidentiality protection of individual data. The Federal Statistical Office of Germany is always working on optimising the balance between these two goods following the special regulations in the statistics act. For international data access this is much more difficult, looking at the different legal situation.

NETHERLANDS [Kooiman]:

Paragraphs 27(ii); 67 & 68:

Balancing the right to know and the need for privacy is difficult since it compares issues of a quite different order. Having researchers analyse data and publishing their findings is a kind of practical, utilitarian benefit, whereas a breach of confidentiallity, or a harmful disclosure is something close to negating and therebye undermining the fundamentals of social behaviour. So the common phrase of just balancing the two is perhaps too simple: it neglects the fact that we have to balance apples and pears which is basically impossible. I would prefer to say that in any case NSO's should behave as decent organisations, keeping their promises to their data providers at any cost. Statistical microdata can never be used for other purposes than intended and communicated to respondents when being collected. This is a frst principle about which we should never negotiate. Nothing to balance whatsoever! Balancing only starts when informed consent is there and NSO's are mandated by respondents to act as their principal agent, who judges case by case whether the benefit of the research outweighs the extra diclosure risk involved.

*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?*

SLOVENIA:

Last year we organized a discussion on this matter with the research community for the first time at the conference in Radenci. For the purpose of giving you the opinion of the Slovenian research community we have met with the representatives of ten institutions dealing with research in different ways (ministries, faculties and research institutes). Our intention was to present them the CES paper and our plan of dissemination of micro data to the research community and receive their view on the subject. From the discussion held at this meeting we can point out a set of conclusions.

It was very clearly pointed out from our research community that data should be protected but that we have to find out the ways and procedures to provide the research community with data to enable secondary analyses within the legislation on data protection. They (and also we at the statistical office) see sharing the responsibility for privacy protection between the national statistical office and researchers as a good and possibly successful principle. Representatives of the research community have seen no particular problems with cooperation in the proposed way.

UNITED STATES:

While both confidentiality and microdata access are public goods, it is not clear how much "balancing" can be achieved. As noted, there may be legal changes required. Nonetheless, confidentiality cannot be treated simply as equally important as access.

NETHERLANDS [Stoop]:

Yes.

NETHERLANDS [Kooiman]:

line 4: unidentifyable --> identifiable

NORWAY:

Note:

Statistics Norway has not undertaken any survey among the relevant research institutions about their views on the issues raised in the Discussion Paper. However, many of these issues were discussed in the report from a commission on "Infrastructure in the Social Sciences" from 2003. Among others, Statistics Norway and research institutions that are extensive users of microdata were represented in this commission. Our comments to the specific points below (Paragraphs 68, 69, 71, 75(b)) largely refer to the conclusions in the report mentioned above.

NORWAY:

There is a general acceptance in the research community that privacy and confidentiality protection are fundamental, not only for the NSO, but also for the research community. It is understood that breach of confidentiality resulting in loss of trust in the protection of confidentiality, will hamper data quality also for research purposes, and may also result in less access to microdata for research.

NETHERLANDS [den Butter]:

Data collected and produced by NSO's indeed have the character of a public good. It implies that the data are, in principle, non-rival and non-excludable. Therefore the guidelines should be clear about the argumentation of excluding potential analysts from using the data. It is, e.g., not obvious that the outcome of the analysis should again have the character of a public good and be available for everybody. As a matter of fact aggregated statistics, such as cyclical indicators and stock market prices, are used by commercial analysts. Why prevent the use of microdata for these purposes when they really are a public good (and the analysts guarantee to respect the guidelines)?

**NOTE:** For comments on paragraphs 69-75 see:

| | |
|---|---|
| Paragraph 69: | Section VI |
| Paragraph 70: | Paragraphs 46-47 |
| Paragraph 71: | Paragraph 53 |
| Paragraph 72: | Paragraph 62 |
| Paragraph 73: | Section XIV |
| Paragraph 74: | Section XIV |
| Paragraph 75(a): | Paragraph 34 |
| Paragraph 75(b): | Paragraph 49 |
| Paragraph 75(c): | Paragraph 53 |
| Paragraph 75(d): | Paragraph 16 |

*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.*

GERMANY:

A question about the respondent perspective should generally be asked in the very beginning of every discussion about access to official microdata. The respondents' opinion is the most important point of view in data confidentiality questions and nevertheless unfortunately hardly ever mentioned. Respondents of the official statistics are the society. They are not only asked, but also answered a lot of questions by the official statistics and also by research of official microdata for societal issues. So the opinion of the respondent and the image he has of official statistics is the benchmark for quality of statistical data of any kind.

Of course there should be efforts to consult the respondent. Especially for business data a kind of waiver-contract could solve some confidentiality problems as mentioned above. But also the individual or household respondent should be asked for permission for the use of his or her microdata. That would be also some kind of risk management in giving back some of the responsibility to the respondent. In Germany the respondent is alluded to scientific access to official microdata on the front page of every questionnaire given out by the German FSO.

UNITED STATES:

As we have also noted, we need more research on respondent perspectives.

AUSTRIA:

We are not familiar with any studies.

NETHERLANDS [Stoop]:

75. (e) Confidentiality concerns do not seem to be a major reason for non-response. The feeling that their participation is important and that the data will be put to good use seems important to respondents.

## ANNEX 1: GOOD PRACTICES

NETHERLANDS [Stoop]:

Good practices : An excellent example outside official statistics, is the European Social Survey, a survey on values, opinions, preferences and behavioural correlates. Data have been collected in 23 European countries according to strict central specifications by universities, commercial organizations and NSIs. Measures have been taken to protect confidentiality. At the same time, the data from the survey are freely available for researchers all over the world. As a result the data, which has been available since September 2003, has thousands of registered users, and hundreds of users have downloaded the data.
www.europeansocialsurvey.org

UNITED STATES:

**National Center for Health Statistics – An Example of Good Practice**

The National Center for Health Statistics is committed to sharing data with researchers regardless of location, within the confines of our legislative mandate and confidentiality requirements. In almost every case, NCHS survey data are collected with the assurance to the respondent that the data will be used for "statistical purposes only" and will not be provided in such a way that the respondent can be identified. The legislation allows for the collection of data for other uses IF the respondent agrees to the use; however, NCHS rarely uses that prerogative.

Much of the data collected from NCHS surveys is available in Public Use files. However, these are not summarized files. They are microdata files with a separate record for each respondent. Many of these files can be downloaded via the Internet for no cost; others can be obtained on disk. Although these data are anonymnised microdata files, NCHS decided to have the users read and agree to a statement that advises that no attempt will be made to identify any respondent (see Attachment 1).

Of course, not all levels of detail can be provided in the public use format. Information on small geographic areas, individual ages and races, and details about facilities/offices in which patients are seen, etc., are often withheld due to confidentiality concerns. The Center has developed a Research Data Center (RDC) to assist in providing more detailed information without compromising confidentiality. Other data files can, on occasion, be linked to the NCHS file also. Two access methods are available for researchers to accomplish their tasks: Onsite and remote access. Work is continuing on making the RDC as useful as possible to the researcher. Costs and time have been concerns. Significant time is devoted to reviewing and approving the research that is proposed. More detailed information about the RDC is in Attachment 2.

Currently the Center is developing a sworn agent process, part of the new Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA). When approved this process may be used in selected situations to allow researchers the opportunity to work at NCHS on their approved research activities.

NCHS has collaborated with other countries in the health statistics arena for many years. Issues related to access to care, insurance coverage, and hospital utilization by type of facility are some of many topics that can be researched using public use data, but analyzed in more detail using the non-public use data.

**U.S. NCHS PUBLIC USE/ MICRODATA USE RESTRICTION NOTICE**

(The information below is found on the "Read Me" section of data available on the Internet as well as on disks.)

!WARNING -- DATA USE RESTRICTIONS!
READ CAREFULLY BEFORE USING

The Public Health Service Act (Section 308 (d)) provides that the data collected by the National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention (CDC), may be used only for the purpose of health statistical reporting and analysis. Any effort to determine the identity of any reported case is prohibited by this law.

NCHS does all it can to assure that the identity of data subjects cannot be disclosed. All direct identifiers, as well as any characteristics that might lead to identification, are omitted from the dataset. Any intentional identification or disclosure of a person or establishment violates the assurances of confidentiality given to the providers of the information. Therefore, users will:

1. Use the data in this dataset for statistical reporting and analysis only.
2. Make no use of the identity of any person or establishment discovered inadvertently and advise the Director, NCHS, of any such discovery.
3. Not link this dataset with individually identifiable data from other NCHS or non-NCHS datasets.

BY USING THESE DATA, YOU SIGNIFY YOUR AGREEMENT TO COMPLY WITH THE ABOVE-STATED STATUTORILY-BASED AGREEMENTS.

**U.S. NCHS Research Data Center**

The continuing demand for analyses that require restricted data with lower levels of geography such as States, counties, and smaller areas, but without confidential identifiers such as names or social security numbers, has been the impetus for the creation of the Research Data Center (RDC) located at the National Center for Health Statistics (NCHS) headquarters in Hyattsville, Maryland. Designed for the researcher outside of NCHS, the RDC allows access to data that would not be permissible to analyze because of confidentiality/disclosure rules and regulations. Information that would, if accessed with no restrictions whatsoever, be considered identifiable and not releasable can under the restricted conditions of the RDC be subject to statistical manipulation. While information concerning named geographic entities cannot be accessed, data ordered by such units can be analyzed at a level not possible with public use data.

Prospective researchers must submit a research proposal that will be reviewed and approved by a committee whose judgment is based upon the availability of RDC resources, consistency with the mission of NCHS, general scientific soundness, and the feasibility of the project. It is expected that the user will develop the research proposal with the RDC staff to minimize the time required. Although researchers will sign confidentiality agreements, strict confidentiality protocols require that researchers with approved projects must complete their work using the facilities located within the RDC. Researchers can supply their own data to be merged with NCHS data sets. Completed by the RDC staff, the merged files will be only available to the originating researcher unless written permission is given to allow access to others.

Two access methods are available for researchers to accomplish their tasks: Onsite and remote.

Onsite researchers have the ability to use the full capabilities of the SAS system with the only caveat being a disclosure review. However, major restrictions limit the analysis capabilities of the remote access system. Totally scanned and screened, printed output using SAS procedures and functions adhere to strict minimal disclosure limits and are suppressed if the minimums are not met.

Another analysis option for researchers with large, complex analytic projects may be to subcontract with the NCHS RDC to perform the tasks necessary for the research project.

Regardless of the option chosen, each has an associated cost that includes capital and labor expenses involved in setting up and monitoring each project. Cost estimations both in terms of time and money, can only be as accurate as the degree of specificity of the research. Again,

with the help of the staff of the RDC at an early stage, the estimates should reasonably approximate reality.

## ANNEX 2 GENERAL APPROACH AND INSTRUMENTS AT EU LEVEL

EUROSTAT:

At EU level, statistical confidentiality is addressed in a series of instruments which are the following:

a. Council Regulation (EEC, Euratom) No 1588/90 of 11 June 1990 on the transmission of data subject to statistical confidentiality to the Statistical Office of the European Communities;
b. Council Regulation (EC) No 322/97 of 17 February 1997 on Community statistics;
c. Commision Decision 97/281/EC of 21 April 1997 on the role of Eurostat as regards the production of Community statistics;
d. Commission Regulation (EC) No 831/2002 of 17 May 2002 implementing Council Regulation (EC) No 322/97 on Community Statistics, concerning access to confidential data for scientific purposes.

The release concerns primarily EU research institutes. Other institutions (ECB, OECD, scientific organisations outside EU, …) are subject to explicit agreement on their admissibility for carrying scientific research with confidential micro data. The current criteria for admissibility take into account the aim of the organisation, the scientific reputation, the organisation of research within the institute, the organisational provision regarding data security, the use and dissemination of the results.

Currently, the access to micro data available at Eurostat is considered through the following channels:

a. The release of EU harmonised and anonymised micro data sets
b. The safe center (data laboratory) located in Eurostat premises. However, the infrastructure in place is limited and this possibility should only considered for very few special cases.

All micro data transmitted by MSs to Eurostat are to be considered as confidential. Public use files are not allowed by the current framework. The access to confidential micro data files for research is conditional to NSO's agreement on the specific research work to be carried out.

The organisational burden implied by the current legal framework is important and has so forth prevented Eurostat to develop it in full extent.