WP. 48
ENGLISH ONLY

UNITED NATIONS STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE CONFERENCE OF EUROPEAN STATISTICIANS

EUROPEAN COMMISSION STATISTICAL OFFICE OF THE EUROPEAN COMMUNITIES (EUROSTAT)

Joint UNECE/Eurostat work session on statistical data confidentiality
(Geneva, Switzerland, 9-11 November 2005)

Topic (vii) General statistical confidentiality issues

MANAGING STATISTICAL CONFIDENTIALITY AND MICRODATA ACCESS – DRAFT PRINCIPLES AND GUIDELINES OF GOOD PRACTICE

Invited Paper

Note prepared by Dennis Trewin, Australian Bureau of Statistics
On behald of the Task Force on Confidentiality and Microdata 1)

1) This draft was prepred and circulated as a Room Document as background for discussion at Agenda Item 6 for the 53rd Plenary Session of the Conference of European Statisticians.

UNITED NATIONS **E**

Economic and Social Council

Distr.
GENERAL

CES/2005/5
7 June 2005

ENGLISH ONLY

STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE

CONFERENCE OF EUROPEAN STATISTICIANS

Fifty-third plenary session
(Geneva, 13-15 June 2005)

MANAGING STATISTICAL CONFIDENTIALITY AND MICRODATA ACCESS – CORE PRINCIPLES AND GOOD PRACTICE GUIDELINES

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

This draft is still under development. It has not yet incorporated all the comments received on earlier drafts. It is being circulated as a Room Document as background for discussion at Agenda Item 6 for the 53rd Plenary Session of the Conference of European Statisticians.

One section that does require further work is that dealing with international access. Specific views on this section will be sought at the CES.

Further comments would be most welcome and should be sent to dennis.trewin@abs.gov.au or tiina.luige@unece.org.

MANAGING STATISTICAL CONFIDENTIALITY AND MICRODATA ACCESS - PRINCIPLES AND GUIDELINES OF GOOD PRACTICE

INTRODUCTION/BACKGROUND

1. Historically, confidentiality protection has been mainly a national issue. However, in the context of increasing data dissemination over the Internet, it is now becoming also an international issue. There is a great deal of international collaboration among members of the research community, and the researchers can be very critical towards different access rules in different countries. Furthermore, researchers are often not allowed to access other countries microdata because of the fear that confidentiality protection cannot be guaranteed. Nevertheless cross country comparisons can be a very important part of a research project. This is not only of interest to the academic community. International agencies are among those who want to use microdata for research purposes, particularly cross country comparisons.

2. This raises the question of whether it is possible to internationally agree on some common principles for dissemination of microdata. This question should be seen in the context that the 2003 Conference of European Statisticians (CES) agreed that support for research is an important activity of the National Statistical Offices (NSOs), and generally NSOs could do more to satisfy these needs. Doing more includes providing access to microdata which is the main focus of these principles and guidelines. (Although the reference is to NSOs in these guidelines, in many countries, particularly those with decentralised systems, there are several statistical producers. The reference to NSOs should be read as incorporating all producers of official statistics.)

3. There are two key objectives in these guidelines;

(i) to foster greater uniformity of approach by countries whilst facilitating better access to microdata by the research community; and

(ii) through these guidelines, and supporting case studies, to enable countries to improve their arrangements for providing access to microdata.

4. The term microdata is used throughout the paper. It can refer to data about an individual person, household, business or other entity. It may be data directly collected by the NSO or obtained from other sources, such as administrative sources.

5. These guidelines recognise that the precise arrangements for access to microdata will vary from country to country. They will vary depending on matters such as legislation, public attitudes and the capacity to support the research community. For example, the arrangements for a well developed statistical office will be quite different from those in a less well developed statistical office. It should not be anticipated that each country will come up with precisely the same arrangements, although it is hoped these guidelines will lead to greater uniformity of approach.

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

7. A number of countries have existing legislation. Specifically the European Union

(EU) has legislation on confidentiality which embodies several principles and rules. These will already be applied by many ECE countries, especially the EU countries. It is recognised that existing legislation is not easily changed and that changes to existing guidelines require collaboration with a range of stakeholders. But opportunities do arise from time to time and these guidelines may be useful in determining appropriate changes. Indeed, in some countries, these guidelines may provide a useful stimulus for debating and agreeing on changes.

8. Any questions on these guidelines should be submitted to the Statistics Division of the UNECE.

WHY SHOULD NATIONAL STATISTICAL OFFICES SUPPORT THE RESEARCH COMMUNITY?

9. In most countries, official statistics are collected not just for governments but for the use of the community. This is particularly the case in democracies where official statistics can be used to assess the effectiveness of government's policies and programs - they provide a mirror on society.

10. To quote a 1993 White Paper on Open Government in the United Kingdom.

> "Official statistics are collected by government to inform debate, decision making and research both within government and by the wider community.
>
> "They provide an objective perspective of the changes taking place in national life and allow comparisons between periods of time and geographical areas.
>
> "Open access to official statistics provides the citizen with more than a picture of society. It offers a window on the work and performance of government itself, showing the scale of government activity in every area of public policy and allowing the impact of public policies and actions to be assessed."

11. The research community plays a particularly important role in stimulating policy analysis and debate and assessing the effectiveness of government programs. This requires access to good quality statistical data if their analyses are to be effective. If they do not have access to relevant official statistical data, they will often seek to collect their own data. As well as incurring additional costs to both the data collector and the respondent, these collections will often be of lower quality.

12. Thus, providing researcher access to microdata can be a way of extracting valuable insights into the quality of the data and how statistical surveys might be improved or extended.

13. Thus, providing researcher access to microdata can be a way of extracting additional value from the cost of collecting official statistics.

14. What is the research community? It includes those working in academic institutions of course, both within a country or outside the country. It also includes researchers working

in non-government organisations and international agencies. Indeed some researchers requiring access to microdata will work within government funded agencies and institutions. For the purposes of these guidelines they are also regarded as part of the "research community" as the microdata confidentiality issues are just as relevant to this group. As will be seen from these Guidelines, the pertinent issues vary somewhat between the different elements of the research community.

15. The following section tries to bring together the perspectives of national statistical offices and the research community in the spirit of trying to find arrangements that largely satisfy the needs of both groups. This latter point is considered in more detail in Chapter 6.

The perspective of the National Statistical Office

16. 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.

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

18. 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.

19. 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 (eg 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 substantially to these costs.)

20. 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 are used more extensively in this way, it can provide an extra level of protection to budget reductions to these statistical programs.

Nevertheless, NSOs are the custodians of data which has been trusted to them and they are responsible for the legality, consistency and transparency of practice.

The perspective of the research community

21. From the perspective of the research community, supporting research based on microdata should be an important component of any official statistical system. The benefits include the following.

(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. For example, the US Bureau of the Census has formalised the documentation it requires from researchers to assist it to improve the quality of its surveys.

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

22. Furthermore, 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 compared with researchers using different data sets to analyse particular topics.

23. 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.

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

> "Perhaps through this additional concept of utility, people will recognise that while they surely have the right to privacy, they may also come to the realisation that they

have a duty to share information, if the common good is to be furthered."

Some use the term "privacy deficit" recognising that there are privacy issues associated with microdata release. The discussion can then focus on whether the benefits of a proposal outweigh any privacy deficit.

25. 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, 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.

CORE PRINCIPLES

26. The Sixth UN Fundamental Principle is very clear on statistical confidentiality.

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

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

27. The following principles should be used for managing the confidentiality of microdata. Each is discussed in the following paragraphs.

Principle 1: It is appropriate for microdata collected for official statistical purposes to be used for statistical analysis to support research as long as confidentiality is protected.

Principle 2: Microdata should only be made available for statistical purposes.

Principle 3: Provision of microdata should be consistent with legal and other necessary arrangements which ensure that confidentiality of the released microdata is protected.

Principle 4: The processes for researcher access to microdata as well as the uses and users of microdata should be transparent, and publicly available.

28. Making available microdata for research is not in contradiction with the sixth UN Fundamental Principle as long as it is not possible to identify data referring to an individual. Principle 1 does not constitute an obligation to provide microdata. The National Statistical Office should be the one to decide whether to provide microdata or not. There may be other concerns (for example, quality) which make it inappropriate to provide access to microdata. Or there may be specific persons, institutions or purposes when it would be inappropriate to provide microdata.

29. For Principle 2 a distinction has to be made between statistical and administrative purposes and uses. In the case of statistical use, the aim is to derive statistics that refers to a

group (be it of persons or legal entities). In the case of administrative use, the aim is to derive information about a particular person or legal entity to make a decision which may bring benefit or harm to the individual. For example, some requests for data may be legal (a court order) but inconsistent with this principle. It is in the interest of public confidence in the official statistical system that these requests are systematically refused. In order to assess whether the intended use of microdata is consistent with statistical purposes, a "compatibility" test can be applied. If the use of the microdata is incompatible with statistical purposes, then microdata access should not be provided. Ethics committees or a similar arrangement may assist in situations where there is discretion whether to provide access or not.

30. The term statistical purposes should be explained. Of course researchers are accessing microdata for research purposes but to support these analysis they may make need to compile statistical aggregations of various forms, compile statistical distributions, or to fit statistical models, or to analyse statistical differences between sub-populations. These uses would be consistent with statistical purposes. To the extent that this is how the microdata is being used, it could also be said to support research purposes.

31. With respect to Principle 3, legal arrangements to protect confidentiality should be in place before any microdata can be released. However, the legal arrangements have to be complemented with administrative and technical measures to regulate the access to microdata and to ensure that individual data can not be disclosed. Existence and visibility of such arrangements is necessary to increase public confidence that microdata will be used appropriately. The arrangements should also be cleared with the privacy authorities of countries where they exist before they are established by law. If such authorities do not exist, there may be NGOs who have a "watchdog" role on privacy matters. It would be sensible to get their support for any legal or other arrangements or at least address any serious concerns they might have.

32. Principle 4 is important to increase public confidence that microdata is being used appropriately and to show to researchers that decisions about microdata release are taken on an objective basis. While it is up to the National Statistical Office to decide whether, how and to whom microdata can be released, the principles of organising the access and taking the decisions to grant or not to grant the access have to be transparent. The NSO web site is an effective way of compliance and also for providing information on how to access research reports based on released microdata.

SUPPORTING LEGISLATION

33. Legislation supporting microdata release is very important as highlighted by Principle 3 (see Chapter 3). There are several reasons.

(i) to provide public confidence in the arrangements - that there are legal constraints that determine what can and cannot be done;

(ii) to provide mutual understanding between NSOs and researchers on the arrangements;

(iii) to provide for greater consistency in the way research proposals are treated; and

(iv) to provide a basis for dealing with breaches.

34. If legislation is not available, some other form of authorisation is essential. The reputation of the NSO is at some risk if there is not some form of authority to enable the release of microdata even when anonymised.

35. It is important that the legislation (or authorisation) cover the following aspects:

(i) what can and cannot be done and for what purposes;

(ii) the conditions of release; and

(iii) the consequences if these conditions are breached.

36. Case Study 1, 2 and 3 outline the legislation for the provision of microdata for Australia, United States and Finland respectively.

METHODS OF SUPPORTING THE RESEARCH COMMUNITY

37. There are various ways a National Statistical Office can support research work. These are summarised below. There is more expansive commentary in the following paragraphs. Case studies are used to further illustrate these different methods.

A. Statistical products for use outside the National Statistical Office

Dissemination Stream	Notes
(i) Statistical Tables/Data Cubes	This can include both standard tables and special tables (or special analyses for that matter) generated at the request of the researcher. Some offices now release very detailed matrices, known as data cubes, which researchers can manipulate to support their own needs. However, if very detailed, the level of confidentiality risk can be similar to microdata.
(ii) Anonymised Microdata Files (AMFs) - public use files	These are microdata files that are disseminated for general public use outside the NSO. They have been anonymised and are often released on a medium such as CD ROM sometimes through a data archive. (Note: The term anonymised implies that, not only are names and addresses removed, but other steps are taken to ensure that identification of individuals is highly unlikely.) Public Use Files are generally available. The level of confidentiality protection in Public Use Files should be such that licensing is not necessary. Public Use files have been a common way of providing access to researchers in many countries.
(iii) Anonymised Microdata Files - licensed files	Licensed files are distinct from public use files in that use may be restricted to approved researchers and a legal undertaking is signed before files are provided to them. They may contain potentially identifiable data particularly if linked with other data files.

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

Service	Notes
Remote Access Facilities (RAF)	Arrangements are now being made in many countries that allow researchers the ability to produce statistical outputs from microdata files through computer networks and without the researchers actually "seeing" the microdata. Because of the additional controls that are available through RAF, and the fact that microdata do not actually leave the NSO, access to more detailed microdata can be provided this way.

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

Service	Notes
(i) Data Laboratories (DL)	On-site access to more identifiable microdata, typically with stringent audit trails and NSO supervision. The access to more detailed data creates some inconvenience to the researcher, because of the requirement of working at the NSO, or at an NSO enclave.

Statistical tables/data cubes

38. Statistical Tables remain the most economical way of satisfying many research needs. Their importance should not be underestimated. The advent of data cubes (very detailed multi-dimensional tables) has increased the usefulness of statistical tables for research purposes as they allow researchers to manipulate the data cubes to suit their own needs.

39. Statistics Netherlands was one of the early organisations to embrace data cubes. Case Study 4 illustrates how they have developed data cubes.

40. Confidentiality issues still exist for statistical tables and data cubes. For example, most statistical legislation requires that identifiable data cannot be released through statistical tables. But the "confidentialisation" is done prior to release. Software systems exist for confidentialising statistical tables and improved methods continue to be developed. They are often referred to as Disclosure Avoidance methods.

Anonymised microdata files - public use files

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

42. Although NSOs generally provide equality of access to all users of their statistics, this may not be appropriate for microdata. A different attitude may be taken to users who do not have strong bona fide research credentials or if they have access to data bases where it would be easy to match AMFs.

43. 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.)

44. Users have emphasised 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, particularly for countries with smaller populations and those with population registers. Prior to the release of Public Use Files there should be 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. Generally, the level of risk will be much greater for countries with smaller populations. Consequently, researchers should not expect that all countries will release Public Use Files.

45. Case Studies 5 and 6 describe the arrangements for the release of Public Use files in United States and United Kingdom respectively. It is interesting to note the role that Social Data Archives play in managing access to individual researchers.

46. There is extensive literature available on the methods for anonymising microdata files. A good summary is available in Willenburg, L & de Waal, T (2001), "Elements of Statistical Disclosure Control. The software package, m-ARGUS, is concerned with protection of microdata against disclosure. Several techniques are available in m-ARGUS.

Anonymised microdata files - licensed files

47. This is an arrangement where specific users are authorised or licensed to use an anonymised microdata files. There will be conditions associated with the licence. These may vary country to country or even from one researcher to another depending on the research proposal and possibly the affiliation of the researcher.

48. The conditions may include some or all of the following

- an agreement by the researcher that he or she will abide by the conditions of release:
- no attempt will be made to identify particular persons or organisations;
- the information will only be used for statistical or research purposes;
- the microdata will not be provided to other persons;
- the microdata will be returned to the NSO when the research project is completed; and
- no attempt will be made to statistically match with other data bases without permission.

49. It is good practice for such an undertaking to have some legal standing, i.e. incorporated within enabling legislation. This would allow legal actions to be taken in respect of breaches. This does not preclude other actions that might be taken in respect of breaches such as not providing any further services to the researcher and/or possibly the researcher's institution. These are discussed in Chapter 7.

50. It should be possible to release more data through licensed files than public use files if some reliance can be put on the undertaking to ensure protection of the confidentiality of the data. That is, some of the data is potentially identifiable, especially when linked other files. It is not good practice to release microdata files when spontaneous identification (eg of well

known people) is possible. NSOs should retain discretion over the approval of users and uses of microdata. Ethics Committees may be able to assist NSO heads in making decisions where discretion has to be exercised.

51. Case studies 7 and 8 describe the arrangements for the release of licensed microdata files in Australia and the Netherlands respectively.

Remote access facilities

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

53. In reality, there are two basic 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. The acceptability of different arrangements will vary country by country.

54. 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.

55. From the cost perspective, RAFs are preferable to Data Laboratories (see below) as the supervised access in a RAF is less expensive than the supervised use involved in Data Laboratories.

56. If these facilities do not remove identification risk entirely, there should still be some

agreement made by researchers to ensure they are fully aware of their obligations. It is good practice to only provide access to those researchers who have signed some form of agreement outlining the conditions of access. Education is also important, together with regular monitoring and checking of the use of these facilities.

57. Case studies 9, 10 and 11 outline the remote access facilities in place in Canada, Australia and Denmark respectively.

Data laboratories

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

59. Some NSOs have established new premises for data laboratories in locations that are more convenient to researchers, (sometimes known as Research Data Centres) but this also can be an expensive option (unless specific funding is provided to the National Statistical Office).

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

61. Case studies 12, 13, 14 and 15 outline the data laboratory arrangements in Canada, USA, Netherlands and New Zealand respectively.

Engaging a researcher as a temporary NSO staff member

62. Another way that researchers may access microdata is through engaging as a temporary NSO staff member and being subject to the same secrecy provisions as the staff of the NSO. This should really be seen as a work program issue rather than a data access solution. It should not be done unless the researcher is assisting with the work of the NSO - otherwise it could be seen as a sham. If this type of pretence was occurring and became public, confidence in the NSO would diminish.

63. The involvement of the researcher may be at the initiation of the NSO - they are seen as someone who can bring special skills to the work of the NSO. On the other hand the proposal may come at the initiation of the researcher. But the NSO must accept the merit of the proposal and incorporation in its work program activities. It is easier to demonstrate that they were assisting the NSO if a published NSO output resulted from this work.

Business data

64. Another special issue is Business Data including agricultural businesses. These are 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 some countries, data bases of business data are often more accessible thereby enabling matching. In addition, many academic researchers might also serve as consultants to business and even bona fide access to business microdata by them might be incompatible with such consultant roles (they cannot be brainwashed of knowledge acquired in the course of their research). Moreover, countries may have economic competitiveness (and possibly even security) issues with respect to sharing identifiable business data with researchers in other countries.

65. From the point of view of researcher access, the main differences between household/personal data and business data are that the dissemination streams that provide greatest protection are most relevant to business data.

66. In terms of the dissemination streams

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

- Anonymised Microdata Files may only be relevant for the smallest businesses. For some research work, this 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). They are likely to be of limited use.

- 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. But, if large businesses are included, it may be difficult to confidentialise outputs even if the researchers cannot directly access the microdata.

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

MANAGING THE TENSION BETWEEN NATIONAL STATISTICAL OFFICES AND RESEARCHERS

How might the tension between the NSO and researcher perspectives be resolved?

67. 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.

68. There are risks that have to be managed. The rapid expansion of data bases, 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. (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 a medium sized country like Australia, about 25% of households are unique at the national level 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.)

69. 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. 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 unfortunate incident.

70. Transparency is important to avoid accusations of secrecy. Therefore, it is good practice for NSOs to be 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. These discussions have to be managed carefully or the privacy fundamentalists could sway public opinion. Support from respected and authoritative persons is very important.

How do NSOs manage the risks?

71. Some suggestions are outlined below.

(i) Agree on a set of principles (such as those outlined later in Chapter 3) which should 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 with 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 on the merits of the research proposal and the credibility of the researcher. Regardless, NSOs must conform with the legislation or other protocols that operate in their country.

(iii) To have an arms length process for the balancing of these two public goods. Ethics Committees may be able to assist in situations where there is discretion in deciding whether to provide access or not. The public good arguments are much stronger if the research is to be placed in the public domain.

(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 not be possible without considerable "distortion" of the data. Explore other opportunities to use technological developments 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 understand the reasons NSOs are so protective of confidentiality. Ensure researchers are aware of the consequences to them and their institution if there are breaches. Follow through on retribution if there are breaches.

72. The last point requires some comment. The culture and value system of the research community is very different to that which operates in an NSO. They often regard some of the "controls" inherent in the microdata access arrangements as unnecessary bureaucracy. Whilst there are no known incidents of researchers using their access to microdata to identify individuals, there have been incidents where microdata provided to them on an exclusive basis being provided to other researchers without permission, or of cases where microdata has been statistically matched without permission with other data to produce richer data sets. The researchers in question may feel they have done nothing wrong as they have not tried to identify individuals but incidents of this type can undermine public confidence. NSOs and researchers operate in different cultures and take different views of risks from incidents. This has to be taken into account in the determination of procedures for release of microdata.

How can NSOs put some of the risk back on to researchers?

73. This might include:

(i) Asking them to prove their bona fides as a researcher. Demonstrating the public benefits of their research and the microdata is "fit for purpose".

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

(iii) Explain the reasons NSOs are being cautious. Ensuring researchers are fully aware of their obligations through appropriate training. Follow-up with effective audit and monitoring procedures. It may be useful to establish a Code of Conduct in collaboration with the research community.

(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 appropriate disciplinary action against the offender). Undertaking legal action where appropriate.

74. 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 should be at the discretion of the NSO. There will be responsibilities associated with access. In particular, researchers should accept that they will have a shared 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 there is a good reason and they still must be followed.

75. It is sometimes argued that respondent consent should be sought before release of microdata outside of the NSO. This should be discouraged.

(i) There are significant practical issues associated with seeking and managing consent.

(ii) Data being provided is unidentifiable and is only being used for statistical purposes, consistent with the purpose of the data collection.

(iii) It is very difficult to provide all the information required for a respondent to make an informed decision - hence many respondents will say No just as a precaution. The sample will soon become unrepresentative.

However, there is an obligation, as stated elsewhere in these guidelines, to be transparent about the arrangements. By this means, in countries where it is a legal option to not seek specific consent, it can be argued that passive consent has been obtained.

(Note: Informed consent would be appropriate in a situation where the publication of small aggregates allows users to infer the situation of a single unit that is part of this aggregate. This situation is more likely to apply to business statistics.)

76. There is another perspective of consent. The data of a NSO comprises data collected directly by themselves and data collected by administering authorities and passed on to the NSO. Unless there is specific provision in legislation to the contrary, a NSO should not release data from administrative sources without the consent of the administering authority (who may feel unable to give consent because of promises made to their respondents). Even when administrative data is in the public domain, it would be courteous to advise the administering authority to give them an opportunity to comment.

77. 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?

(i) 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.

(ii) 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.

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

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

MANAGEMENT ISSUES ASSOCIATED WITH THE RELEASE OF MICRODATA

Managing decision making on confidentiality

78. It is not possible to be prescriptive about whether identification is possible or not in a microdata release. There is always some chance of identification, even when very small. Software now exists which can estimate the proportion of records which are unique and therefore at some risk of identification.

79. It is the chief statistician or his or her delegate who needs to make the decision on the release of a microdata file whether it be by an anonymised microdata file (public use or licensed) through a remote access facility, or through a data laboratory. In order to make that decision, the chief statistician needs advice on whether, for example

- the risk of identification is sufficiently small;
- the adjustments made to the data items have not unduly damaged the microdata file for research purposes; and
- the variables that have been collapsed are the most appropriate taking into account both the needs of researchers and identification risk.

80. As an illustration of the last point, choices could be made between the amount of detail provided on geography and household structure.

81. Appropriate arrangements should be put in place to provide this advice on a consistent basis. It often needs to be supported by a research capability and could be located in a methodology area. Case studies 16 and 17 describe the arrangements in Slovenia and Australia respectively.

Managing meta data

82. If users are to make effective use of microdata, they must have access to the appropriate meta data. This would include:

(i) a description of survey including any information on quality;

(ii) a list of the data items and the classifications used (sometimes referred to as a Data Dictionary); and

(iii) definitions of the data items.

Provision of (i) will help ensure that the microdata are not used inappropriately.

83. As microdata are provided electronically, the meta data must be provided in a way that is accessible. Printed copy may still be the most effective means.

Managing decision making on individual research requests

84. Again this is the responsibility of the chief statistician although the criteria for making the decision are quite different.

85. The three key criteria are:

- whether the research satisfies ethical considerations; and

- perceptions of the benefits of research can be enhanced if the proposed research benefits are confirmed by qualified peer review and if the results are put into the public domain.

There will also be resourcing and opportunity costs to consider.

86. Again, arrangements to provide advice to the chief statistician are necessary. These will be different to the arrangements mentioned in paragraph 4. For example, it may be useful to set up an Ethics Committee as part of these arrangements. External involvement is more likely in these arrangements.

Managing breaches by the researcher

87. Efforts should be made to reduce the likelihood of breaches as outlined in the previous chapter. Nevertheless, breaches may occur and procedures for dealing with the breaches should be determined.

88. Breaches must be treated seriously. If this is not done, public confidence in the arrangements will erode. Also, breaches are more likely to occur if they are not treated seriously.

89. There are a number of ways of dealing with breaches. For example, if a legal offence has occurred, legal action should be considered. This is expensive but is essential to demonstrate the importance the NSO places on confidentiality, and reduce the likelihood of future offences.

90. Also, the researcher should be prevented from further access to microdata. This should be the minimum step that is undertaken.

91. Consideration should also be given to stopping further release to the institution of the researcher, at least until,

(i) the institution has taken appropriate steps in dealing with the offence committed by the researcher; and

(ii) the NSO is confident that the appropriate arrangements are in place within the institution to minimise the chance of further breaches.

92. The research community should generally be supportive of taking strong action against the relatively small number of offenders who may give the research community a bad name. It is in their long term interest.

SOME SPECIAL ISSUES

International access (requires further discussion)

93. 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 there are also risks. For their work, staff from international agencies are not subject to any national or international legislation other than the applicable staff rules of the organisation. Some care has to be taken. The main difficulty is that the scope for retribution against breaches is much more limited for researchers living in other countries. But, on the other hand, the probability of identification is much less. A further difficulty is that some countries do not have the legal authority to provide data to researchers outside their country.

94. How can researchers access data sets from other countries? How can international agencies obtain access to microdata for statistical and research purposes? The options include:

(i) Public Use Files where they exist,

(ii) Licensed Anonymised Microdata Files,

(iii) Remote Access Facilities with appropriate safeguards.

Some international agencies may already have arrangements in place for some statistical enquiries and it is not necessary to consider these options.

95. Public Use Files are only available for some countries. Licensed anonymised microdata files may be an option if enabled by the legislation of the NSO, but also might depend on the trust in the researcher and his or her institution. This is a choice of the NSO but for many would 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. An exception may be made for Eurostat where specific legislation and rules have been established to protect the confidentiality of microdata provided by member countries.
105. For many countries, the use of Remote Access Facilities may be the preferred route to provide access to international researchers in the medium term. Under such arrangements,

there are more controls and the position of NSOs, on international access to microdata, is more easily defended if challenged. However, the usability of these arrangements in this environment still needs to be developed.

96. Experience has shown that there are some legitimate areas of concern in providing access to international researchers. These include:

- passing on the microdata to other researchers without authority; and

- assuming ownership of the microdata and not returning it on requests from the NSO (e.g., it may contain significant errors).

For these types of reasons, many NSOs are cautious.

97. NSOs will need to decide whether they can provide access to international researchers or not taking into consideration the range of issues discussed in this section, keeping in mind that legal penalties can only be applied according to the law of the country where the violation has been committed. They should keep in mind that a risk management approach is being encouraged. For some research applications, the benefits may justify the risks involved as long as the arrangement is legal. For some institutions, the risks may be lower than for others. NSOs will also need to decide the most appropriate form of access. International researchers, including international agencies, will have to accept that RAF may be the only possibility for some countries.

98. There is another possibility. International researchers, including the international agencies, could work through networks of national researchers to achieve their aims. Indeed, these national researchers could be located in the NSO for international studies of particular importance.

99. Case Study 18 describes the arrangements that are in place for the Luxemburg Income Study.

Data linking

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

101. 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.

102. While, there are clear benefits in data linking, there are also risks, particularly if the source agency for the linked file is not the NSO. 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 Chapter 3 are followed for linked data sets.

103. For those countries with Privacy Commissions or like bodies, the arrangements for data linking should be supported by the Privacy Commission.

104. Case Studies 19 and 20 describe how Canada and Sweden respectively manage data linking arrangements in their countries.

ANNEX 1.1

CASE STUDY - LEGISLATION TO SUPPORT RELEASE OF MICRODATA - AUSTRALIA

1. Broad description

This is legislation which enables the National Statistics Office to release microdata to approved users for statistical purposes. It also outlines the conditions of release and the penalties for any breach of those conditions.

2. Why is it good practice?

It provides a degree of certainty to both the National Statistical Office and the potential users of microdata about the arrangements for release. The legislation also outlines the arrangements that the Parliament is happy with. As they are enshrined in legislation, they are in the public domain.

3. Target audience

Primarily the research community who are the main users of microdata.

4. Detailed description

The specific legislation is outlined in Part 5. There is also a supporting statement providing policy, rules and guidelines to assist ABS staff involved in the release of microdata.

Each new release of microdata requires the approval of the Australian Statistician in view of the potential sensitivity of releases. Each release to individual clients requires the approval of a senior manager, employing the delegated authority of the Australian Statistician.

A Microdata Review Panel has been established to provide advice to the Australian Statistician on microdata releases, particularly the steps that need to be taken to ensure the release complies with the confidentiality test imposed by the legislation.

5. Supporting legislation

Disclosure of unidentified information

(1) Information in the form of individual statistical records may, with the approval in writing of the Statistician, be disclosed if:

(a) all identifying information such as name and address has been removed;
(b) the information is disclosed in a manner that is not likely to enable the identification of the particular person or organisation to which it relates; and
(c) the Statistician has been given a relevant undertaking by each person required by sub clause (2) to give a relevant undertaking in relation to the information.

(2) The persons required to give a relevant undertaking are:

(a) for information to be disclosed to an individual — the individual; and
(b) for information to be disclosed to an official body:
(i) the responsible Minister in relation to, or a responsible officer of, the official body; and
(ii) if the Statistician considers it necessary in a particular case — each individual in the official body who will have access to the information; and

(c) for information to be disclosed to an organisation other than an official body:
(i) a responsible officer of the organisation; and
(ii) if the Statistician considers it necessary in a particular case — each individual in the organisation who will have access to the information.

(3) In this clause:

relevant undertaking means an undertaking in writing that use of the information in relation to which the undertaking is given is subject to the following conditions:
(a) no attempt will be made to identify particular persons or organisations to which the information relates;
(b) the information will be used only for statistical purposes;
(c) for information to be disclosed to an individual — the information will not be disclosed to anyone without the approval in writing of the Statistician;
(d) for information to be disclosed to an official body or other organisation:
(i) the information will not be disclosed to anyone outside the body or organisation without the approval in writing of the Statistician; and
(ii) if the Statistician considers it necessary in a particular case — the information will not be disclosed to an individual in the body or organisation who has not given a relevant undertaking;

(e) if the Statistician considers it necessary in a particular case — either or both of the following:
(i) the information, and all copies (if any) of the information, will be returned to the Statistician as soon as the statistical purposes for which it was disclosed have been achieved;
(ii) access by officers to information, documents or premises will be given as may be necessary for the purpose of conducting a compliance audit concerning observance of the conditions under which the information is disclosed;

(f) any other condition that, in the opinion of the Statistician, is reasonably necessary in a particular case.

In a different part of statistics legislation, it is made clear that a person who fails to comply with an undertaking, as prescribed in (2) above, is guilty of an indictable offence publishable on conviction by a fine not exceeding $5000 or imprisonment for a period not exceeding 2 years, or both.

6. Strengths

(i) Provides a basis for arrangements that are understandable by both the National Statistical Office and researchers.

(ii) Provides for significant penalties for legal breaches - may be a reason why no known breaches have occurred.

(iii) Microdata protection is partly provided by a legally enforceable Undertaking. This means that some protection (eg prevention of matching) can be provided through the Undertaking.

(iv) Provides wider access to the data than would otherwise be the case, thereby achieving a greater return on the high investment in data collection and respondent burden.

(v) Provides statutory authority and transparency for release practices and a basis for the public defence of those practices

7. Weaknesses

(i) Researchers still believe the conditions of release are too limiting ie the steps taken to make identification not likely result in too much of the detail not being released.

(ii) Disclosure of detailed information, even under circumstances demanding strict confidentiality, can alarm the privacy constituency and in the worst case, have potential to impact on response rates?

(iii) Compliance with the limitations and conditions imposed by legislation can impose an administrative burden on both the NSO and the users, delay the release of information, restrict the range of users who can have access to the information, restrict the uses to which the information can be put and limit the nature of the information which can be released.

8. References

The supporting statement referred to in Part 4 is available on request from teresa.dickinson@abs.gov.au

ANNEX 1.7

CASE STUDY - ANONYMISED MICRODATA FILES (LICENSED FILES) - AUSTRALIA

1. Broad description

Anonymised Microdata Files (Licensed Files) in Australia are referred to as Confidentialised Unit Record Files (CURF). Key measures undertaken by the ABS to protect the data are: requiring anyone who uses the data, and the organisations that employ them, to sign an undertaking with the ABS; obtaining user commitment to confidentiality principles; and perturbing data or reducing detail on files to make it very difficult for units to be identified. CURFs are most commonly made available to users either on a CD-ROM or through a remote access data laboratory (RADL™). CURFs available on RADL™ contain more detail than those on CD_ROM. In selected cases users may have access to a CURF through an on-site data laboratory.

2. Why is it good practice?

Releasing microdata in this manner constitutes good practice as:
perturbation of data and masking of records is undertaken to maintain the integrity of the data while protecting the confidentiality of an individual's data; and
placing restrictions on how the data are used, as set out in a legal undertaking to be signed by each user and their organisation, ensures both the user and the organisation accept responsibility for keeping the data confidential and secure.

3. Target audience

CURFs are aimed at Australian researchers and analysts within government, academia and other non-government organisations, who seek to undertake more in-depth analysis than is possible using tabular aggregated data.
CURFs are not generally released to overseas applicants. In very selected instances the ABS allows overseas researchers to access CURFs via the RADL™ if they are sponsored by a suitable Australian organisation. The sponsoring organisation is required to sign an undertaking with the ABS.

4. Detailed description

The ABS has adopted a manner of release for CURFs that protects the data in three ways; confidentialise the unit record file to control the detail available; provide modes of access appropriate to the level of detail available; and require users and organisations with access to the data to sign an undertaking that restricts how they use the data.
The unit record files are confidentialised by removing name and address information, by controlling and limiting the amount of detail available, and by perturbing or deleting data where it is likely to enable identification of individuals.
Each CURF release is personally approved by the Australian Statistician, following advice from a Microdata Review Panel consisting of three senior executives. The panel makes a

detailed assessment of each CURF to ensure that the disclosure risk is low.
There is protection inherent in the different access modes and in the different levels of data provided in each. The ABS provides three different modes of access for CURFs - CD-ROM, the RADL™ and the ABS Data Laboratory (ABSDL). CURFs available on CD-ROM are labelled Basic CURFs and are restricted to a relatively small number of variables released in broad categories. RADL™ users can also access Expanded CURFs that contain more variables and more detail, with extra protection provided by the automatic logging of RADL activity and subsequent audits of this activity. Specialist CURFs contain the most variables and detail and can only be accessed via on-site ABS Data Laboratories.
Each user must apply to be granted access to a CURF, explaining their intended use of the CURF. Both the User and a Responsible Officer of the employing organisation must sign a legal undertaking in which they agree:

access to information about individuals will be restricted to officers of the organisation who have signed an individual undertaking with the ABS;
users will not attempt to identify individuals;
users will not match the unit data to other files of unit data;
ABS officers are allowed access as necessary to audit compliance with these rules;
CURF usage is limited to the specified and approved individual 'Statistical Purpose'; and
any sensitive printed data and output will be stored in a secure place.

The organisation must monitor its officers that have access to the CURF and ensure that all have signed an individual undertaking with the ABS. Access to CURFs are for statistical purposes within an organisation. If an individual changes organisations, they must surrender access and notify the ABS.
The responsible officer is generally the head of an organisation, department or university. They are required to sign an undertaking about the storage and use of the CURF. Breaches can be addressed by sanctions against both the individual user and the organisation as well, including removal of access to all microdata for all individuals in the organisation.

5. Supporting legislation

The release of microdata by the ABS is governed by legislation; namely, the Census and Statistics Act 1905. This legislation enables the Australian Statistician to release unit record data, provided this is done "in a manner that is not likely to enable the identification of a particular person or organisation to which it relates."

6. Strengths

(i) Allows for a range of access mechanisms to suit a range of uses.

(ii) Allows for access to more detailed data to be granted to users who are able to work with a greater level of environmental protections.

(iii) Microdata protection is partly provided by a legally enforceable Undertaking. This means that some protection (eg prevention of matching) can be provided through the Undertaking.

(iv) Sanctions can be applied against users and organisations that breach the undertakings, providing additional motivation to ensure data access and use is appropriate.

7. Weaknesses

(i) Researchers still believe the protections applied directly to the microdata are too limiting. They believe too much of the detail is not being released, especially for some of the most identifiable sections of the population (eg large households).

(ii) It is more costly to support a range of access mechanisms than a single access mechanism.

8. References

The Census and Statistics Act 1905 -
<http://scaletext.law.gov.au/html/pasteact/1/580/top.htm>
The Statistics Determination 1983 -
<http://scaletext.law.gov.au/html/pastereg/0/414/top.htm>
CURF undertakings & the Responsible Access to ABS CURFs Training Manual -
<http://www.abs.gov.au/websitedbs/D3110129.NSF/85255e31005a1918852558ac00697645/72d92417a0ba71b5ca256d01002c47a4!OpenDocument#Untitled%20Section_6>

ANNEX 1.11

CASE STUDY 7 - REMOTE ACCESS FACILITY (FOR MICRODATA ACCESS) - AUSTRALIA

1. Broad description

The Remote Access Data LaboratoryTM (RADLTM) is a web-based tool that allows authorised users to access detailed microdata that is stored within the ABS secure environment. Built-in automatic checks prevent large scale release of unit record information, thus maintaining confidentiality of data providers as outlined in Australian legislation.

2. Why is it good practice?

The RADLTM provides access to more detailed and less confidentialised microdata than can be made available on CD-ROM. It provides greater flexibility in user analysis of microdata.

Access is limited to authorised users. All microdata remains within the ABS computing system. A balanced mix of automatic and manual processes prevent clients from obtaining outputs containing large amounts of unit record information. A justifiable audit trail is automatically maintained.

3. Target audience

The RADLTM is primarily targeted at Australian government agencies involved in policy development and research areas within Australian universities. To a lessor extent, the RADLTM is also used for research purposes by the private sector and by non-profit institutions.

4. Detailed description

Potential users of microdata are required to sign legal undertakings and read training material provided before RADLTM access will be granted. Authorised users are required to comply with published data security guidelines and any further instructions of the ABS.

The RADLTM operates as a three stage process. Clients submit batch-style queries via a secure section of the ABS website, which are firstly parsed for illegal commands. If the query is accepted, it is then executed in conjunction with ABS confidentialised microdata files. Finally, all produced output is automatically checked for confidentiality issues before being made available to clients on a secure web page.

A retrospective auditing process manually checks for inappropriate use of ABS microdata, and provides empirical evidence that automatic checks have been applied appropriately.

5. Supporting legislation

The release of microdata by the ABS is governed by legislation; namely, the Census and Statistics Act 1905. This legislation enables the Australian Statistician to release unit record data, provided this is done "in a manner that is not likely to enable the identification of a particular person or organisation to which it relates."

6. Strengths

(i) Provides a secure on-line access point, from which users may access detailed ABS microdata from their own computing environments;

(ii) Automatic protection of output at time of execution allows quick turnaround;

(iii) Enables the ABS to release more detailed microdata than that which can be released on CD-ROM;

(iv) Flexibility of user analysis. Users are not restricted to a set of predefined tables.

(v) Users are alleviated of CD-ROM security and data storage concerns;

(vi) Statistical software is provided by the ABS. Users do not need to supply their own licenses.

7. Weaknesses

(i) Researchers still believe the conditions of release are too limiting (ie) the steps taken to make identification not likely result in too little detail being released.

(ii) Limited to batch-mode style of programming, lack of graphical user interface functionality.

(iii) Time taken to build automatic protections limits variety of statistical software packages made available.

(iv) Heavy manual auditing load.

8. References

(i) Australian Bureau of Statistics, (2005), Responsible Access to ABS Confidentialised Unit Record Files (CURFs) Training manual, Edition 2, Canberra, Australia, also available at <www.abs.gov.au->services we provide->curfs>

(ii) Australian Bureau of Statistics, (2004), The Remote Access Data Laboratory (RADL) User Guide, Revised Version 2.0, Canberra, Australia, also available at

<www.abs.gov.au->services we provide->curfs>.

(iii) Access to ABS CURFs web pages. <www.abs.gov.au->services we provide->curfs>.

ANNEX 2

STANDARD TERMINOLOGY

National Statistical Office (NSO)

Although the term is used in the singular, it is meant to incorporate all statistical agencies, or statistical units within Government Departments, who produce official statistics and provide access to microdata for statistical or research purposes.

Research community

Although this mainly refers to people working in research institutions such as universities, it also includes researchers working in government agencies, NGOs, international agencies and the private sector. Some countries may want to define the research community more narrowly and only include those working in research institutions.

Microdata

This refers to data about individual persons, households or legal entities.

Statistical purposes

It is particularly important to make a distinction between statistical and administrative uses. In the case of statistical use, individual data are used as an input to derive statistics that refer to a group of persons or legal entities. It may also incorporate support for other activities within a NSO (eg sample selection off a business register). Administrative uses concern decisions about a particular person or legal entity which may bring benefit or harm to the individual.

The statistics referred to above include statistical aggregates, statistical distributions, parameters for models and other forms of statistical analysis that may refer to groups of individuals or organisations without identifying them.

Anonymised microdata files - public use files

These are microdata files that are disseminated for general public use. They have been anonymised and are often released on a medium such as CD ROM sometimes through a data archive. The term anonymised implies that, not only names and addresses removed, but other steps taken to ensure that identification of individuals is highly unlikely.

Anonymised microdata files - licensed files

Licensed files are distinct from public use files in that use is restricted to approved researchers for approved purposes. A legal undertaking is signed before files are provided to them.

Remote access facilities

These are facilities that provide researchers with the ability to produce statistical outputs from microdata through computer networks without researchers actually "seeing" the microdata. The microdata itself does not leave the National Statistical Office. Remote Access Facilities may be of two types.

(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.

Data laboratories

This involves working on-site at the National Statistical Office, or one of its Branches, to obtain access to microdata. Access could be direct or indirect through staff of the National Statistical Offices. If access is direct, the researcher is in effect being treated as a temporary employee of the National Statistical Office with the inherent responsibilities.

Risk avoidance

This approach tries to eliminate all risks. In the case of microdata confidentiality, it requires the confidentiality of the data to be absolute, not only in its own right, but in association with other available data.

Risk management

Within the constraints provided by legislation, it involves identification of the risks and managing them in accordance with their significance (impact) and their likelihood. More effort is putting in managing the high impact, strong likelihood risks. Microdata confidentiality may not be absolute when considered in association with other data. Confidentiality could be considered in association with other means of reducing the risk.

Data linking

Data can be linked by exact matches (eg using an identifier such as name and address or ID number) or by statistical matches (using probabilistic matches). They may be NSO data sets only, a NSO and administrative data sets, or administrative data sets only. Data sets for a particular collection could be linked longitudinally. All these possibilities are incorporated within Data Linking.

ANNEX 3

ACKNOWLEDGEMENTS

This work is mostly the result of the efforts of a Task Force set up by the Conference of European Statisticians. The Task Force is Dennis Trewin (Australia), Ivan Fellegi (Canada), Otto Andersen (Denmark), Teimuraz Beridze (Georgia), Luigi Biggeri (Italy) and Tadeusz Toczynski (Poland).
Mr Trewin was Chairman of the Task Force.

Svante Oberg also provided considerable assistance to the Task Force during the course of their work.

Several countries provided case studies to support the Guidelines is greatly appreciated.

* * * * *