

SYNTHESIS OF RESPONSES TO THE OECD QUESTIONNAIRE ON MICRO-DATA ACCESS

Draft

Nadim Ahmad, Statistics Directorate, OECD.

Introduction

1. The necessity of micro-data for decision makers is becoming increasingly self-evident. A fact that can be witnessed by the explosion in recent years of papers and conferences that have, at their heart, micro-data, and by the increase in policy measures based on the recommendations and conclusions of micro-data analyses. In response to these developments, in late 2005, the OECD launched a study to investigate the feasibility of the OECD accessing micro-data from statistics institutions and to develop a set of governance principles that could facilitate this, whilst, also, by providing a forum by way of a Conference and a list of recommendations, allow countries to share best practice in the art of micro-data access.

2. The ultimate objective of the study therefore is to provide OECD analysts with a mechanism to access comparable international micro-data to assist policy formulation in Member Countries, but also, to investigate whether the micro-data already accessed by the OECD could be better, and more readily, accessed via an OECD Centre for micro-data; with its own and common governance rules that provided confidence to data providers in the OECD's ability to manage micro-data access and, in-turn, increased the potential for the OECD to access new forms of micro-data.

3. To avoid misunderstanding however, it is important to be clear, from the start, that access should not necessarily be interpreted as acquisition. It is possible, for example, for the OECD to gain access to micro-data through analytical software or associated remote access e-mail based systems, say, that mean the analyst never sees the micro-data. And indeed this is one of the main options being investigated in the study.

4. A full description of the options investigated within the feasibility study are provided as an Annex to this paper, which is the Questionnaire sent to statistics institutes earlier this year that sought, as a first-step, to gain views on the feasibility of each of the options directly from each of the institutions that would potentially be involved. For convenience however the options can be summarised as follows:

- i. Non-disclosive **indicators** based on snap-shots of micro-data.
- ii. Non-disclosive **indicators** based on longitudinal micro-data.
- iii. **Bilateral projects**
- iv. **Remote access** – where a number of possible approaches exist.
- v. **Provision of micro-data** - for storage: short, medium or long-term within an OECD Centre.

5. A further option described, in the annex, is to determine whether access to private-sector data sources provided a more feasible form of access; an area that is not covered in this paper but remains of interest to the study and will be developed over coming months.

Synthesis

Summary

6. The Questionnaire sent by the OECD complements the Questionnaire sent by UNECE in autumn 2005 (Annex B) to statistics institutes to investigate national practices vis-à-vis the provision of micro-data to the international organisations and agencies. The synthesis below, both in this summary and in the detailed country responses that follow, tries to reflect the common position reflected in the responses to the OECD and UNECE Questionnaires. The summary of the UNECE returns, prepared by the UNECE Secretariat is also presented as an annex (C) for convenience.

Access is increasing....

7. Twenty seven OECD and two large non-OECD countries responded to the OECD Questionnaire, and the common message is a positive one. Although the access mechanisms and the data that can be accessed vary considerably across countries, in all but two some form of access to micro-data is made available. This is a good sign as it reflects not only the growing importance of micro-data but also the growing willingness of statistics institutions to meet user needs in this regard.

.....as are the methods of access; remote access solutions in particular

8. Typically, countries make access available to household survey data with a variety of different controls applied to preserve confidentiality, and which vary depending on the dataset. Or rather, which vary depending on the degree to which an individual respondent can be identified. Most countries, for example, make public-use files available, and many others go one step further and provide access to anonymised files whose use is governed by a license or contract stipulating the terms of use. This latter type of file is typically made accessible within a research centre. But a not insignificant proportion of countries have begun to move towards more accommodative access methods, recognising the geographical constraints and impracticalities imposed on researchers having to visit a research centre, and these countries have begun to or are beginning to experiment with remote access solutions.

Typically household data are made accessible but the legal constraints are usually no different for business data

9. Admittedly most of these developments have taken/are taking place in the area of household micro-data but there are some positive implications here nonetheless for business micro-data, since, in all but two countries (Portugal and the United Kingdom), the regulations governing micro-data access are de facto the same for both household and business data. In the UK, business data are protected by national legislation whereas personal data are protected by the pledge given to respondents under the data protection act, but, importantly, despite these differences, the UK is one of the few countries who are currently able to provide access to their business micro-data. Moreover in Portugal access to household micro-data is in fact more difficult than business data, it being illegal in Portugal to disclose personal micro-data.

The difference in access between household and business micro-data reflects the difficulty in preventing secondary and complementary disclosure, not legal constraints per se

10. In theory, from a legal perspective, therefore, there is very little to differentiate between business and household micro-data access. In practice however the accessibility between the two different types of datasets are very different. Indeed, at the time of writing, only a handful of countries currently permit some form of access to business micro-data. What distinguishes the relatively widespread accessibility to household micro-data from access to business micro-data are the practical difficulties inherent in preserving confidentiality of individual businesses.

11. Suppressing the most obvious identification characteristics, such as name and exact address, of an individual respondent, is a relatively robust and simple mechanism to preserve anonymity and confidentiality as far as household surveys are concerned, so long, of course, as the number of other variables in their response are not too expansive or require too much specification: for example, there's only one Queen in the United Kingdom. Assuming that she were asked to respond to a survey that asked for her occupation it wouldn't take a genius to work out who exactly was responding even if you were to remove just her name and address. But, even in these cases, countries employ a variety of methods to overcome such problems, for example by suppressing such variables or individual variable responses or by grouping variables together until the number of respondents in a particular variable group is below a certain threshold that removes or reduces the possibility of identification, or, by providing only samples of the full micro-data set (all the while ensuring that individuals, and their variables, in the sample data set are drawn from populations above a certain threshold).

12. For business micro-data, particularly in small countries such as Ireland and New Zealand, this process is much more challenging and it is, fundamentally, the difficulty inherent in this process that has so far led to only an, albeit encouraging, handful of countries to begin making such data accessible. But the important point to note here is that it is the inherent difficulty and the risks associated with disclosure that have so far stopped some countries from moving forward in this domain and not the legal constraints per se, since these same countries have been able to make household micro-data available; the availability of which is governed by, de facto, the same legal constraints. Indeed the response by some countries to the OECD and UNECE Questionnaires reflected a view that it was impossible to be able to guarantee non-disclosure in business micro-data sets whilst still retaining the essential characteristics (detail) of the micro-data itself.

13. The fact that this is so is not contentious. Most users recognise that business data and its analyses usually imply a focus on a, usually small, sub-group of businesses, be they high-growth, low growth, high productivity, dominant, multinational, etc. Although not mentioned in the country returns it is possible that complementary disclosure may be part of the reason for prudence in the area of business micro-data. Cigrang and Schouten¹, for example, recognised that it would be very difficult to introduce mechanisms that could prevent complementary disclosure from multiple data requests for example; although, that said, these comments also apply to household micro-data.

But many countries have been able to tackle business micro-data access difficulties by identifying and creating, a trusted, responsible, educated user network

14. But clearly some countries have been able to make inroads in making business micro-data accessible and the methods and procedures implemented by these countries could be of enormous benefit to those who have yet to make this journey. In most cases, the mechanism used to preserve confidentiality

¹ Cigrang, Mark and Barry Schouten (2003) 'Remote Access Systems for Statistical Analysis of Microdata' Methods and Informatics Department, Statistics Netherlands.

is trust, and, so, the onus is placed firmly on users of micro-data. Of course, trust is reinforced by a number of measures:

- *Screening users*: typically, by ensuring that they are researchers from a respected academic or government institution who have been able to demonstrate a strong need and business case for requiring access.
- *Training users*: some institutions, for example the UK Office for National Statistics, provide comprehensive training programmes for users before they are provided with access; programmes that included detailed guidance on how to prevent accidental disclosure and the legal requirements placed on them.
- *Penalties*: legislation in many countries permits severe penalties, including fines and imprisonment, on those users who disclose individual respondents in the micro-data. But in practice the most severe threat for users is that they and their institutions will have access to the micro-data withdrawn, and, often, this includes a threat to 'name and shame'. Given that, in most countries, users and institutions have been previously required to demonstrate a strong need for micro-data, withdrawal of access is usually seen by users themselves as the most severe of sanctions.

And these measures have been successful

15. To gauge the effectiveness of these measures, one need only to count the number of known deliberate breaches that have occurred to-date across those countries where business micro-data has been made accessible; that number being zero.

16. Some countries that make micro-data accessible do so only to residents on the grounds that criminal penalties can only be effectively applied to residents but, although impossible to state categorically, this is perhaps being over-cautious, as the experiences of those countries that rely on the threat of withdrawal of access (to screened and trained users), suggests that this is the most effective and preventative sanction. Indeed one has to recognise that these sanctions will nearly always imply a severe financial cost on those institutions that have access withdrawn; for example: research projects may have to be abandoned mid-way; research grants returned; further research grants jeopardised etc. For an organisation such as the OECD, which has a significant reliance on voluntary contributions, withdrawal of access rights could be catastrophic both financially and in terms of its reputation.

But many other options can provide simple solutions to user needs in the immediate to short-term

17. But that all being said it is perhaps important to think a little outside of the box, or, at least to reinforce the notion that micro-data access should be viewed as a process, where the box is being increasingly and gradually opened. The first two of the OECD options highlighted above, for example, imply no breaches in disclosure because, by definition, they merely request that statistics institutions think a little more creatively about the types of standard indicators that can be created where confidentiality can be guaranteed.

For example, by expanding the scope of standard structural business statistics....

18. A good example of this process reflects the now standard dissemination by statistics institutes of structural business statistics data. It was not so long ago that such statistics were not commonly disseminated, indeed their primary purpose, was to serve as an input into the national accounts or to assist government industrial policy. But, in response to user requests for more detailed information on industry output, turnover, employment, profits etc, these statistics are now widely disseminated with well

established rules that prevent disclosure. Clearly in some, particularly small, countries these rules mean that the data on some industries are suppressed and not made available but users recognise and understand the reasons for this. However it is possible to go considerably further in this direction by providing new indicators that continue to respond to user needs whilst all the while preserving confidentiality. For example, many statistics institutes² could, in theory, readily provide the following set of indicators to complement their structural business statistics, disaggregated by 2, 3 or 4 digit industry groupings and perhaps size class:

- (Hirschman-Herfindhal) concentration indices by 2, 3 or 4 digit industry groupings;
- Labour productivity, 95% confidence intervals by 2,3 4 digit groupings and size class; and arguably maximum and minimums too;
- Labour/Capital ratios, 95% confidence intervals by 2,3 4 digit groupings and size class; and arguably maximum and minimums too;
- Business Demography statistics, on birth, death and survival rates (an area that has already seen significant expansion particularly in EU countries), high growth enterprise rates, gazelle rates.
- Standard regression coefficients, for example linking output to labour and capital using KLEMS' type production functions say, that allow for the possibility of differentiating between different types of businesses; in particular institutional status and ownership structure.

19. Clearly not all of these indices could be produced for all sectoral/size-class breakdowns. Concentration indices for a sector with only 2 or 3 businesses would not, for example, guarantee non-disclosure. But suppressing these cases would still mean that a significant set of standard indicators could still be provided to the great benefit of users.

... or, by capitalising on new developments in disclosure control techniques

20. The indicator approach however provides but one mechanism to the un-locking of micro-data access. Many other possibilities exist. For example a number of statistics institutes are experimenting with disclosure control methods such as synthetic data, perturbed data, etc and these are certainly developments that could be shared across institutes, although, of course, with some implications on resources.

But, ultimately, the best mechanism, for users, is for access to anonymised data, as long as statistics institutes take the necessary measures to create a network of educated and responsible users

21. The downside of these types of approaches, however, is that they pre-define the scope of analyses for users. It will never be possible, for example, to create a perturbed/synthetic data set/list of indicators that can respond to every possible user requirement, and, as such it will always be preferable to provide users with access to the actual dataset that allows them to conduct their own analysis. It is important to note however that this does not necessarily require that the users see the actual data. Most, if not all, users would be content with merely being able to run statistical applications on the data, albeit, recognising that they still have a responsibility to ensure that their results remain non-disclosive and it is clear that this approach is already paying dividends in a number of statistical institutes.

² Indeed just such an approach has been undertaken by academics, see for example Bartelsman, Haskel and Martin, 2006, "Distance to Which Frontier?"

In concluding

22. The country returns indicate that statistics institutes are rising to the challenge of making micro-data more accessible. At present such access is typically at the level of household data where it is clearly easier to preserve confidentiality than is that case for business micro-data. But many institutions have begun to address this. Some already make business micro-data accessible and many others are investigating and experimenting with a number of approaches. Two, complementary, approaches, that apply equally to household and business data seem viable and on the basis of the responses to the OECD Questionnaire, most favourable. The first is for statistics institutes to increasingly consider the standard and common types of statistics that are produced by users of their micro-data, and, so develop and disseminate non-disclosive standard indicators.

23. The second is to build on the successful practice of some statistical institutions, based on the notion of a trustworthy and educated user community who are permitted access but with the threat that access for them and their institution will be removed in the case of breaches of confidentiality. It stands to reason that in order to maximise the use of their micro-data statistical institutes should try as far as is possible to extend its reach beyond those analysts resident within the national boundaries of the statistical office, and so provide access using remote access systems. Indeed, it was the second approach that appealed to many of the countries that responded to the OECD Questionnaire as a way of making micro-data accessible to the OECD. A few countries did express a preference for providing access within their national borders at safe centres but this, unfortunately, does not seem a practical solution for the OECD, with a current membership of 30 countries spanning the globe.

24. That said, there has to be recognition, explicit in the returns of countries, that remote access systems do not currently exist in most countries and that these systems would need to be developed over time, and, so, such a solution would need to be seen as a medium term objective. At the same time however, there should be no reason why a remote access system could not be developed in the shorter term with those countries that were currently able to provide access along such lines. Clearly there will be some obstacles to overcome, chief amongst these is the preference, although not requirement, that countries provided access to their micro-data using common formats and variables and concepts.

25. Some countries also indicated that they would be able to provide anonymised micro-data to the OECD for storage, usually for a pre-determined period of time. This is clearly an option but one that does not seem practical in the medium term, as it is not clear that all countries will eventually be able to make micro-data available in such a way, even in the longer term. Remote access solutions are arguably preferable because it should be possible to develop systems where: (a) OECD users need not necessarily see the micro-data; and (b) where returns generated by the system could be automatically screened to ensure that, at the most basic level at least, results remained non-disclosive, for example, using rules that suppressed statistics that were based on a number of observations/individuals below a certain threshold and dominance rules, for example, if 1 or 2 observations were responsible for a significant (threshold determined) proportion of any variable. Of course, if efficient automatic safe-guards could not be implemented to the requirements of some statistical offices it may be necessary to additionally develop remote-execution systems where output files are screened by individual statistics offices before being returned to the OECD.

26. It has to be recognised however that this implies a burden on statistical offices although at the same time the experience of the Luxembourg Income Study, that is able to process and return 100s of requests per day, suggests that the burden may be manageable.

27. The next steps therefore are (i) for the OECD to approach some countries to see whether a pilot remote-access system could be developed³ and (ii) for the OECD to engage with countries and researchers to identify lists of target indicators that could be readily produced in a standard way across countries whilst all the while preserving confidentiality, coupled with the development of Governance rules that provide the basis for data access between the OECD and institutes. Another idea that has not hitherto been given considerable attention is to investigate whether historic data is more easily accessible than more recent data.

³ See Bruno Giuseppe and Nadim Ahmad, “Assessing the feasibility of remote access solutions for micro-data access – a proposed pilot study”; paper prepared for the OECD Conference: Assessing the feasibility of micro-data access, Luxembourg 26/27 October 2006.

Summary Country Responses

Australia

28. The Australian Bureau of Statistics has made a considerable number of micro-data files accessible. Their preferred method of access is using what they call Confidentialised Unit Record Files (CURFs) via the Remote Access Data Laboratory (RADL)⁴. A list of currently available CURFs is shown below. Two types of CURFs are made available, 'Basic' which provides a (broad) level of detail, and 'Expanded' which contain extra data items and more detailed breakdowns. For example a Basic CURF might contain 'Age' with 5 year groupings but on the Expanded CURF, 'Age' would be available as single years.

29. Aspects of Literacy (1996), Basic; Australian Housing Survey (1994) (1999) both Basic; Australians' Employment and Unemployment Patterns (1994-97) Basic & Expanded; Business Longitudinal Survey (1994-98) Basic; Census - Household Sample File (1981) (1991) (1996) Basic; Census - Household Sample File (State/Territory or Section of State) (1986) Basic; Census 2001 - Household Sample File (2001) Basic & Expanded; Child Care Survey (1984) (1993) (1996) Basic; Child Care Survey (1999) Basic & Expanded; Child Care Survey (2002) (2005) Expanded; Domiciliary Services, Victoria (1986) Basic; Disability, Ageing and Carers (1993)(1998) (2001) Basic; Education and Training Survey (1989) (1993) (1997) (2001) (2005) Basic; Education and Work Survey (2001) (2003) (2005) Basic; Families in Australia (1992) Basic; Family Characteristics Survey (2003) Expanded; Forms of Employment (1998) Basic; General Social Survey (2002) Basic & Expanded; Giving to Charities Population Survey Monitor (1997) Basic; Household Expenditure Survey (1975-76) (1984) (1988-89) (1993-94) (1998-99) Basic; Household Expenditure Survey and Survey of Income and Housing (2003-04) Basic & Expanded; Income and Housing Survey (1981-82) Basic; Income Distribution Survey (1986) Basic; Income and Housing Costs Survey (1990) (1994-95) (1995-96) (1996-97) (1997-98) (1999-00) Basic; Income and Housing Costs Survey (2000-01) (2002-03) Basic & Expanded; Labour Force Survey & Employee Earnings, Benefits and Trade Union Membership (August 2004) Basic; Labour Mobility (1984) (1991) (1994) Basic; Mental Health and Wellbeing of Adults, Australia, Second Revised (1997) Basic; Mental Health and Wellbeing of Adults, Western Australia (1997) Basic; Multi Purpose Household Survey (2004-05) Expanded; National Aboriginal and Torres Strait Islander Survey (1994) Basic; National Aboriginal and Torres Strait Islander Social Survey (2002) Expanded; National Crime and Safety Survey (2002) Expanded; National Health Survey (1977-78) (1983) (1989-90) (1995) Basic; National Health Survey (2001) (2004-05) Basic & Expanded; National Health Survey, Indigenous (2001) Expanded; National Nutrition Survey (1995) Basic; Rental Investors' Survey (1993) (1997) Basic; Rental Tenants Survey (1994) Basic; Superannuation and Employment Arrangements Survey (2000) Expanded; Time Use Survey (1992) Basic; Time Use Survey (1997) Basic & Expanded; Voluntary Work Survey (2000) Basic; Women's Safety Survey (1996) Basic.

30. Only two of the available CURFs are longitudinal – Australians' Employment and Unemployment Patterns (1994-97) and the Business Longitudinal Survey (1994-98)⁵ (which is currently, the only ABS economic survey CURF available).

⁴ The ABS RADL is a secure online data query service that clients access via the ABS web site. Authorised users submit queries in SAS, STATA or SPSS via the RADL web interface and the queries are run against the CURFs that are kept within the ABS environment. The results of the queries are checked for confidentiality then made available to the users via their desktops, usually within two to three minutes, depending on the complexity of the query.

⁵ The CURF consists of 9,571 unidentified statistical records containing data on businesses including information on the industry of operation of the business, the age of the business, the number of locations, details on the major

31. In theory the OECD is able to gain access to Australia's CURF files, the pre-conditions being that: such access would need to demonstrate the benefits to Australia; and the existence of a legal undertaking between the OECD and the Australian Director of Statistics, that would also require OECD officials to submit an Individual User Application and Undertaking stating the statistical purposes for which the CURFs will be used.

Austria

32. None of the 6 major datasets identified in the Questionnaire are longitudinal, although there are plans to do so for the business register in 2008. Confidentialised files on the labour force and household surveys as well as EU-SILC can be made available to the OECD for special bilateral projects and via a research centre. Remote access solutions have not been deemed possible but this appears to be more a reflection of current technical constraints than legal barriers. No information however has been provided on the legal constraints/conditions that the OECD would need to satisfy in order to gain access.

Canada

33. Public-Use (confidentialised files) can be provided on: Population Census data, the Labour Force Survey and Household Surveys – both for bilateral purposes and for storage within the OECD. Confidentialised business related data could be accessed via bilateral agreements but it would not be possible to store this information within an OECD Centre. However, anonymised household and business files could be made available, in theory, via remote-access; although at present such a facility does not exist. Reading between the lines it seems likely that the interpretation of remote access here is either (i) a system that allows users to run analyses that return results that do not breach confidentiality or are of the remote-execution variety where results are vetted by Statistics Canada officials before being returned. Statistics Canada is legally prohibited from providing data that could identify an individual (person, business or organisation) and neither Population Census data nor External Trade Flow data are longitudinal

34. Canada ranked indicators (option II) as their preferred way forward in all cases with bilateral projects coming in second and remote access third in all cases. Provision of micro-data to an OECD Centre was the least favoured option.

Czech Republic

35. The Czech return contained no information on access possibilities on business registers or surveys. The Population Census, Labour Force Survey, External Trade Flow and Household Survey datasets are longitudinal, all of which can be made accessible in an anonymised form, with the exception of the ETF dataset, which can be made accessible in a confidentialised form. Remote access solutions are considered legally feasible but no systems currently exist for such access and the mode is considered 'futuristic'. The preference for the Czech Republic is to develop indicators based on micro-data.

Germany

36. The Labour Force Survey is longitudinal. Plans exist to make the business registers, business surveys and ETF datasets longitudinal in coming years. Current legislation permits micro-data access to

decision maker within the business, union membership and employment arrangements; details of family businesses, hours of operation, types of business comparisons made, sources of business advice, major changes in business operations, business intentions, business planning and business improvement programs; the use of business networks, the use of computers and the Internet; innovation and expenditure on research and development; employment details, export details; a range of income and expense items and a derived profit or loss; stocks, assets and liabilities, a breakdown of equity, equity finance details and capital expenditure.

licensed files within Germany but not outside, reflecting the current lack of international data regulations. It does appear possible however, at least in theory, for the OECD to access anonymised data (for all 6 datasets) remotely using Controlled Remote Data Processing (CRDP). CRDP provides a means to access data using SAS, SPSS or STATA. (Like Australia a dummy structural file is provided to allow users to test their applications/syntax beforehand. Programs can then be submitted and then results vetted and processed before being returned to the OECD (assuming they satisfy confidentiality rules) - although for business and ETF data this facility will not be available for a few years yet.

37. Household based micro-data can also be transmitted to an OECD Centre as (Confidentialised) Public Use Files. Anonymised files are also available but these are only currently accessible within Germany or via remote access. In terms of preferences Germany gave the highest (and only) ranking to the remote access option.

Denmark

38. Denmark's response to the Questionnaire was restricted by the fact that access, and disclosure prevention, appears to be controlled via legal (penal) devices. As such, the Danish position is that access can only be permitted to Danish residents or resident institution. Given this constraint Denmark have responded that no micro-data (in any form) could be accessed by or provided to the OECD. That being said however it is possible that the OECD could work together with Denmark to investigate the scope for developing indicators that could overcome these confidentiality problems and, indeed, to see whether rules could be developed and attached to their remote access system that screened and vetted results before being returned to the OECD; even if these rules were overly prudent and provided an unnecessary degree of protection this would be better than no access at all.

Finland

39. No specific rules exist for a specific survey. Access is not typically provided for micro-data covering the whole population (e. g. population census data) or a whole area (e.g. total population of a municipality) and is typically granted to a sample dataset of anonymised data made accessible with a user-license. More information can be found at: http://stat.fi/meta/tietosuoja/kayttolupa_en.html. In summary however it does seem possible for the OECD to gain access to micro-data as Finland does not differentiate between external and domestic applicants. .

France

40. Business register and business survey data is longitudinal. The population census is too but only for a 1% sample. There is interest from the French Customs Department to develop a longitudinal ETF dataset in the future. Certain public-use files on household data are made available through the INSEE website but not for business data, where the INSEE view is that it is impossible to adequately confidentialise micro-data whilst at the same time retaining its essential micro-data characteristics. That said, access requests to anonymised data are considered by a Comité du Secret, and so there is some scope for the OECD to gain access using this channel.

Greece

41. The Labour Force Survey, Business Register, External Trade Flows and Household Expenditure Survey are longitudinal. Greece has responded positively to options III and V for all datasets. They point out that no public use files are available but that access to confidentialised data for research purposes could be granted subject to and in accordance with Articles 5 and 6 of Greek Law (2392/96 c). They have

responded negatively to all remote access options but this is on technical, rather than legal grounds. Greece's preference therefore was to make data accessible to the OECD via provision to an OECD Centre.

Hungary

42. The Labour Force Survey, Business Register, External Trade Flows and Household Expenditure Survey are longitudinal. Business Register data can be provided to the OECD since, in part, this information is not considered secret in Hungary, although turnover and employment data can only be provided by class. Remote access systems seem to be possible from a legal context with the only barriers being technical. Options III and V are also deemed legally possible – for the Population Census this would reflect confidentialised data restricted to a 5 % sample; for ETF data, confidentialised; and, for other household surveys, anonymised micro-data could be made available..

Ireland

43. Section 34 of Ireland's 1993 Statistics Act enables the Statistics Office to provide anonymised micro-data files where the level of anonymisation is determined by the Director General. Further, these datasets are provided to the ISSDA or Irish Social Science Data Archive (see www.ucd.ie/issda/) and which can be released by the ISSDA free of charge to researchers or students for specific bona fide non-commercial research purposes as long as the user signs a contract that imposes strict conditions on the use of the data. Currently anonymised files exist only for household based data, and these are not longitudinal. No business micro-data is provided as the Statistics Institute does not consider it possible to adequately anonymise data. However it is possible for OECD officials to be appointed as employees for the purpose of carrying out specific statistical research projects over a specified period and on-site. Further information on the above arrangements is set out in the Irish Statistics Office Data Protocol (<http://www.cso.ie/releasespublications/CSODataProtocol.htm>). Information on the Code of Practice on Statistical Confidentiality is available at www.cso.ie/aboutus/statistical_confidentiality.htm. Confidential information is transferred in anonymous form to Eurostat for the compilation of aggregate Community statistics under the provisions of Council Regulations (EC) 1588/90 and 322/97. Under further European and Irish legislation, non-anonymised Balance of Payments related statistical micro-data may also be transmitted to the European Central Bank.

44. The technical option of creating a link between OECD and the ISSDA for access to files stored there could be explored, so long as, the outcome included provision for suitable controls that ensured compliance with the terms of use for the household survey micro-data files.

Italy

45. Only Labour Force Survey data are currently, and, partly longitudinal. Anonymised Data can be delivered under the basis of bilateral projects (Option III) as long as these are joint projects. Data delivered to the OECD (under Option V) outside of these arrangements could be made accessible but only in a confidentialised manner, and for business survey information this would be on the basis of what's already made available under EC regulation 831/2002. Overall the Italian preference is for micro-data to be provided under options (I) and (II), as confidentialised indicators.

Japan

46. Japan has taken a strict interpretation of what is legally permissible and have taken the view that no datasets could be made accessible under options (III) (IV) or (V).

Mexico

47. Labour Force and Business Surveys (excluding the Census) are longitudinal. Mexico's clear preference is for indicators defined by the OECD that could be made available and accessed via the INEGI website. What can be done with other access modes is not clear. On balance it would appear that the other options, especially on business data, are not feasible given legal, technical and resource constraints. It is not impossible however that a system that allowed remote access with an interface to preserve confidentiality, such as the mechanism used by Australia and Germany, could be developed.

Netherlands

48. None of the datasets is longitudinal at present. The Dutch have a strong preference for Options (I) and (II), and see these as presenting few problems but at the same time they stress the importance of recognising that any such indicators would need to be clearly specified and may imply additional burdens on statistical offices. Bilateral projects are feasible as long as contracts are signed-up to in advance. Co-ordinated remote-access options are feasible but the Dutch stress the importance of encouraging the harmonisation of country data-sets if this is option is tow work for the OECD. Option (V) is possible for confidentialised Labour Force Survey data only as long as access is restricted to designated OECD users.

Norway

49. Population Census, Business Survey and External Trade Flow data are longitudinal, and Labour Force Survey data can be made so. The overall preference is for indicator work, but this varies depending on the dataset. Confidentialised ETF data for example scores a '1' for remote access but Confidentialised Business Survey data a '1' for provision to a Centre. That said all datasets can be provided concordant with procedures corresponding to Eurostat regulation 02/831) as long as the data is confidentialised.

New Zealand

50. The Business Register and External Trade Flow datasets are longitudinal and the possibility of making Business Survey datasets longitudinal is under investigation. Bilateral projects (option (III)) are considered feasible for access to anonymised Labour Force Survey, Business Survey and External Trade Flow datasets but only on-site and under contract and, moreover, the work would need to be of significant benefit to the New Zealand Official Statistics System before it could be sanctioned. Remote access to confidentialised data, with results screened and vetted before being returned, could be made possible for the Business Register and the Labour Force Survey (in theory, as the remote access system is under development and is not yet available) with an associated cost for access to the Business Register. Remote access could also be provided for aggregate outputs from a confidentialised Population Census file.

51. Option (V) is possible using CURF files, which are currently available for: Population Census data (a heavily confidentialised sample); the 2003 Labour Force Survey; the 2001 Household Savings Survey; and the Income Survey (series of supplements to the Household Labour Force Survey) for 2002, 03, 04. A first CURF is in production for the Household Economic Survey (2003/04). In all cases a contract with the OECD, along with undertakings from all researchers who will use the CURF covering terms and conditions of use for the service, would be required.

52. New Zealand also expressed a concern about the possible negative public perception that might occur if Population Census information was made available to the OECD and the impact this might have on future response rates. Another important point reflected the limitations of what was feasible, particularly in the context of confidentialising business related data, given New Zealand's relatively small economy. In summary New Zealand expressed a preference for indicators fro business related data but remote access methods (using confidentialised files) for household based data.

Poland

53. The Polish return was based on a relatively strict interpretation of what was permissible under Polish law, meaning that the feasibility of Options (III) (IV) and (V) was very limited. Polish law prohibits identification of individuals and specifically (especially) forbids the publication of aggregated data with less than three entries and where the share of any one individual is more than $\frac{3}{4}$ of the total. However household based statistics can be accessed in accordance with Commission Reg 831/2002.

Portugal

54. Only household expenditure and income surveys are longitudinal. All data can be provided in confidentialised form for all of Options (III) (IV) and (V), subject to the 5th article of the National Statistical System Law). A clear preference is for remote access, where all datasets except the Population Census score 1, which scores 1 for direct delivery (Option (V)). Direct delivery scores second highest for all datasets). Access can only be given under a written agreement which should state clearly the obligation of both the OECD and the statistical office and the conditions of use.

Slovakia

55. All 6 data sets are longitudinal. Raw Business Register could be provided as part of a bilateral project and under Option (V). Anonymised data sets could be provided, in theory, for all other datasets as part of Option (V) except business survey/census data, which can be provided on a confidentialised basis (in other words the same business survey data provided to Eurostat before the application of their secondary confidentiality rules, this means all business survey data so long as aggregates do not contain less than 3 entries and an individual with over a 90% share. Remote access solutions are not considered at present technically feasible.

Spain

56. Spain recognised the importance of the OECD desire for micro-data and its importance to economic analysis and policy formulation but they drew attention to their legal code which prohibits access to anything other than public use files. That said they do cite a precedent for Eurostat (Commission Regulation (EC) No 831/2002 concerning access to confidential data for scientific purposes) and point out that the OECD could access this data via Eurostat. The public use files that are available can be found at <http://www.ine.es/prodyser/microdatos.htm>; which consist, in the main, of (human) population and health statistics, including LFS statistics but they also include an ICT survey.

Sweden

57. Citing legal restrictions embodied in their Secrecy Act, Sweden's preference is for the OECD to further investigate the possibilities available under option (IV) - Coordinated remote access - on the basis of the experiences of Australia, Statistics Denmark, and their own system (MONA).

Switzerland

58. Of the 6 datasets referred to in the Questionnaire only the Labour Force Survey and Business Register are currently longitudinal. The Population Census and other Household Surveys are expected to become longitudinal in 2010 and 2015 respectively. All datasets can be provided to the OECD under bilateral agreements supported by a contract that stipulates the OECD need and the period of time for which the data will be made available, after which it must be returned or destroyed. The preferred position of the Swiss was for the development of indicators based on micro-data. All other options for access scored very lowly (6) for options IV, V(i) and V(ii).

Turkey

59. Business Survey and External Trade Flow data are currently longitudinal (with Business Register data expected to be so by the end of this year).

60. Anonymised Labour Force Survey and Household Expenditure Survey micro-data data will be shortly available on request via CD. In the meantime the OECD can request this data Confidentialised Business Register, Business Census and ETF data can also be made available but strictly subject to the rules governing disclosure, and subject to agreements reached with the OECD. Turkish law prohibits the release of confidential data (unless approved by the 'Presidency') where confidential can be defined as: If the number of the statistical units in any aggregation is less than three or one or two of the statistical units are dominant the cell is suppressed. Moreover confidential data compiled, processed and preserved for the production of official statistics cannot be delivered to any administrative, judiciary or military authority or person, and cannot be used for purposes other than statistics or as an instrument of proof and not passed on to third parties.

United Kingdom

61. Confidentialised Population Census and Labour Force Survey data can be made available via a special license, subject to the data being for statistical research purposes. OECD would be eligible to receive the data as it is a 'research organisation' and this is consistent with the confidentiality pledge given to survey respondents. A written agreement setting out organisational and technical measures is required. Business survey data can be accessed on-site, subject to agreements being signed, but other forms of access to business based data are not considered practicable. A remote access option is not impossible but it would require very secure systems and confidentiality protection rules as well as having to overcome the difficulties posed by access from a site outside of the UK intra-government firewall (A possible solution would to have a dedicated server outside of the government net).

United States

62. A non-starter is for confidential data to reside within the OECD due to legal and administrative requirements. Public use files are however available for most household surveys and the population census. In addition, tabulations of census and survey data are available from FedStats (www.fedstats.gov) and specific agency Web sites. The Interuniversity Consortium on Political and Social Research at the University of Michigan (www.icpsr.umich.edu) and the Integrated Public Use Microdata Series at the University of Minnesota (www.ipums.umn.edu) are also useful resources for published statistical datasets. Unfortunately, there are very few public use files available for businesses primarily due to the difficulty in ensuring confidentiality for large businesses.

63. The U.S. Census Bureau is experimenting with remote access technologies that allow users to have analyses run by the agency using the non-public data, with the results returned to the user if they meet the disclosure avoidance requirements. Testing is underway for the Current Population Survey (CPS) and is planned for the American Community Survey (ACS). When the Census Bureau micro-data analysis system is proven to meet its dual requirements to ensure data utility and protect confidentiality, it can be used to generate linear regression coefficients, implement dummy variables and provide a synthetic residual. The Census Bureau expects to implement the system for CPS later this year and for ACS sometime in 2007. However, it should be noted that this system is primarily seen as an access solution for household data. Business data pose additional problems due to the highly skewed distribution of the sample. Nevertheless, some economic census and trade data could potentially be analysed remotely, but the study population and analyses may be restricted. Another access alternative that is under development by the Census Bureau involves creation of inference-valid synthetic data. OECD could use these

disclosable data sets for conducting cross-country studies. Like the micro-data analysis system, however, the utility of inference-valid synthetic data for businesses depends greatly on the skewness of the businesses. The current research effort is being directed in support of the Longitudinal Employer-Household Dynamics program and is designed to provide broader access to the data generated through this program.

Brazil

64. No longitudinal data sets currently exist. All datasets, excluding the Business Register, are accessible on-site and under-contract. No data can be supplied to the OECD under option V but anonymised data (excluding business census or Business Register data) can be provided via remote (internet subscribed) access. The overall preference therefore is for remote access solutions, except for business data, where option (III) is preferred.

Russia.

65. All data, except Population Census data, are longitudinal. Anonymised Labour Force Survey data can be made available however. Business register information is not typically available but it is possible that access could be granted to a confidentialised data-set. That said data on company registrations from the tax office is accessible. External Trade Flow data not sufficiently well developed to permit significant analysis. Social data could be made accessible post 2007, after the development of a system to confidentialise data.

ANNEX A : OECD QUESTIONNAIRE ON MICRO-DATA AVAILABILITY

Introduction

66. Recent policy developments and initiatives⁶ have highlighted the need for better quality and more comprehensive micro-data to provide the tools for evidence based decision making and to tackle complex issues such as understanding innovation, balances and imbalances in economic statistics, and the many implications of globalisation. Developments involving extensive use of micro-data are now occurring in a number of countries and international organisations. Indeed, in many institutions a great deal has been achieved over the past years, using a variety of approaches that has significantly improved the availability of micro-data compared to a few years ago.

67. Through the experiences of these studies, the benefits of micro-data analysis are being extended and made more visible. Some of the most significant work has been in the general area of understanding the influences on firm level productivity and through the associated growth in the significance and benefits of cross country comparisons, further benefits could be attained if micro-level statistics were derived through a more collaborative coordinated approach that ensured greater harmonisation of concepts and statistics.

68. A large share of research studies now carried out by the OECD involve country level micro-data, while studies such as PISA reflect the growth in micro-data collection and management by the OECD itself. The OECD is increasingly a custodian of country micro-data, usually collected explicitly for cross county comparative studies. But these only begin to scratch the surface of the potential for the use of micro-data; many applications are possible, including:

- The measurement of *productivity* and its determinants;
- *Innovation* – and its benefits, identifying the characteristics of innovative firms and people, etc; *ICT, R&D* - and their benefits.
- *Entrepreneurship* – What are the factors of business success? What are the characteristics of high growth firms? etc.
- *Poverty abatement* – What are the key characteristics of people below the poverty line: race, gender, education, region disability etc
- *Personal Wealth and Income Studies* – Which parts of society save most/least? Who is most vulnerable to pension difficulties etc
- *Income inequality* – what are the characteristics of people at different ends of the income scale etc
- *Employment* – Which types of firms create the most employment? Are the self employed wealthier/happier than the employed? What determines this? etc
- *Globalisation* – Are multinationals more efficient, productive than domestically owned firms? Which types of businesses have benefited most from globalisation? What role has outsourcing played? Identification and impact assessment of transfer pricing etc
- *Labour Mobility/Migration* – Is increased labour mobility a good thing? What are the determinants, preconditions for high levels? etc

⁶ E.g. The Bologna Charter on SME Policies and the ‘Bologna Process’; The 2nd OECD Ministerial Conference on SMEs, Istanbul 2004; European Commission Lisbon Summit and the ‘Lisbon Strategy’, 2000.

- *Policy evaluation* – What is the impact of fiscal/social policy on its target audience? Does higher personal/corporate taxation increase fiscal avoidance and by whom? etc.
- *Environmental issues* – Can businesses with higher green credentials prosper? Within similar industries who pollutes most/least? Do local policies play a role? etc

69. The growth in the experience at the OECD with coordinating research projects based on micro-data has up to now generally been on a one-off basis, and it has now reached a level where it is time to give consideration to establishing a more permanent framework. A first proposal for such a framework was presented to the OECD Committee on Statistics (CSTAT) in 2004 [STD/CSTAT(2004)12]. While delegates confirmed the value of micro-data analysis, they were also concerned about costs and legal constraints that a project at the international level would encounter. To gain a better understanding of the options and implications of micro-data analysis at the international level, the Secretary General allocated means from the Central Priority Fund to carry out a feasibility study in 2006. The CSTAT 2005 meeting also recognised the importance of this subject and the coordinating role the OECD could offer and recommended that the OECD organise a conference on micro data in 2006 to exchange best practices and explore the role of the OECD in fostering work with micro data in OECD member countries.

70. The OECD is evaluating the benefits of establishing an OECD Centre for the custody of the micro-data that it continues to draw on for its work across directorates. The Centre could therefore bring together practices and facilities for managing and organising micro-data that is already provided to the OECD. In establishing these facilities and governance rules the Centre could also provide the perfect vehicle to increase its coverage of micro-data, whether that increase in coverage is via remote access processes, where the micro-data is retained within national and international institutions, or whether the micro-data is provided directly to, purchased by, or gathered via specific statistical surveys initiated by the OECD, or brought together by groups of countries when collaborating with the OECD on projects that involve cross country studies based on micro-data. In establishing the Centre, it is expected that the practices and systems, including codes of practice, could be transferred to countries that seek to extend their own micro-data analytical capability, within their own legal and institutional arrangements. The OECD will also look to develop governance arrangements that give countries continued assurance that micro-data held in the custody of, or accessible by, the OECD is done so in a manner consistent with the obligations given to those who provided the information. The Centre would therefore develop into a unique place where the analysis of key policy and statistical issues, requiring international comparisons based on micro-data, could take place to the benefit of the OECD and its Member Countries.

71. There is a continual mix of research projects that involves micro-data of businesses or people, which is an increasingly valuable research resource to the OECD and participating countries. By taking a leadership role in the standards and practices adopted for their management, these resources can be maintained with high confidence. There is a growth in bilateral and multilateral projects where micro-data is shared and compared. The IMF portfolio survey is an excellent example of this. This needs to be facilitated by the establishment of a neutral trusted environment where micro-data can be managed.

72. As a critical first step subsequent to the CSTAT 2005 meeting, the OECD Statistics Directorate presented five options describing roles in which the OECD could fulfil this co-ordination role, to delegates of the Statistics Working Party of the OECD's Committee on Industry and Business Environment Meeting. Delegates requested that the OECD should submit these options to National Statistics Institutes, in the form of a questionnaire, to determine the feasibility, of each of these options and, so, providing an assessment of the best-way forward. This Questionnaire is the response to that request, see also STD/STAT/RD(2006)2, which provides further information on the feasibility study and conference. It builds on earlier work and questionnaires in this field, for example that conducted recently by UNECE following the June 2005 Conference of European Statisticians which concluded that more information was needed on the provision of micro-data to international organisations and researchers from other countries

to better prepare the guidelines on “Managing Statistical Confidentiality and Micro-data Access-Principles and Guidelines of Good Practice.”

73. The report “Guidelines and Core Principles for Managing Statistical Confidentiality and Microdata Access”, prepared by a Task Force set up for the purpose of establishing these guidelines, under the guidance of the CES, and chaired by Dennis Trewin, the Chief Statistician of the Australian Bureau of Statistics, was approved by the CES Bureau in February 2006 and endorsed at the 2006 CES meeting. The full report is available at <http://www.unece.org/stats/documents/ece/ces/2006/6.e.pdf>. The 4 key principles stemming from this report are as follows:

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

Principle 2: Micro-data should only be made available for statistical purposes

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

Principle 4: The procedures for researcher access (including international organisations, government etc) to micro-data as well as the uses and users of micro-data should be transparent and publicly available.

74. This earlier work has provided a wealth of information concerning the legal situation governing micro-data in many countries (43) and their current micro-data access practices and so this questionnaire takes a more restricted focus on the ability of countries to provide micro-data access to the OECD for the use of the OECD Secretariat. Arguably one of the most important conclusions of the CES Task Force however was the fact that in the absence of systems that make micro-data available to researchers, researchers often seek to collect their own data; incurring additional costs to themselves and respondents and results that are typically of lower quality than could be provided by official sources.

75. Questionnaires should be completed and returned via e-mail to Mr Nadim Ahmad at nadim.ahmad@oecd.org **before 28 July 2006**. Any questions concerning the questionnaire should also be addressed to Nadim Ahmad.

Glossary of Terms

Pure data set/raw data – This refers to the full data set with all data and identifiers and, so, where individuals can be identified

Anonymised/ation - This refers to the case where the key identification characteristics, (such as individual business/person name, VAT number, address etc) of micro-level data have been suppressed.

Confidentialised/ing – This refers to the process one-step up from anonymisation which leads to the suppression of micro-data because other variables related to an individual (business/person) can be used to identify the subject with a very high probability. This also refers to the case where aggregates are suppressed because the aggregate information can be used to derive estimates, with a high degree of confidence, related to an individual. Anonymised and Confidentialised data sets differ only in the level of anonymity preservation therefore and are commonly referred to collectively as anonymised micro-data files, whether as public-use or licensed files, which restrict the use of the files to the designated researchers/research institution.

Perturbed/ation – This refers to the process where information on some variables related to an individual is allocated to other individuals, whilst at the same time preserving totals for standard aggregates.

Artificial data – This refers to dummy data that is created purely for the purpose of allowing users to test programs. Typically the programs are then transmitted to the custodians of the data who can then run the programs on the actual datasets and return results to users if they satisfy national disclosure rules.

SECTION 1 - GENERAL INFORMATION

Name of institution reporting to the OECD

«s1a1»

Address of institution

«s1a2»

Website of institution

«s1a3»

Contact Details of Respondent/s (by database, as applicable)

«s1a4»

SECTION 2 – ASSESSING THE FEASIBILITY OF THE OECD OPTIONS

Option Descriptions

Five options have been formulated for this feasibility study, described in turn below. Each component differs in the amount of costs and resources involved, and whether these are on-going (as some will have high start-up costs but lower continuation costs), the level of access to confidential data and the risks of unintentional disclosure. It is important to note that the options are not mutually exclusive. In developing a Centre for micro-data it is already possible at this early stage to envisage the possibility of a multi-track approach, where separate options are pursued for different groupings of countries and data.

The options relate primarily to the different modes of access that can be made and not the data acquisition processes themselves. For example in developing governance rules for the management of micro-data the Centre aims to become not only the place where currently available micro-data sets can be accessed but also the forum that can govern the collection of new micro-data sets commissioned by the OECD and its Member Countries and the access modes to these new data-sets could fall under any of the 5 feasibility options.

Option (I) Decentralised work with longitudinal data

This is the option that has been followed by OECD in previous longitudinal studies (see work by STI and ECO); the OECD plays a role in that it coordinates the research question and basic methodology and synthesizes the results. But micro-data are accessed and treated in a fully decentralized fashion. Proposals for further research along these lines are discussed by SWIC on 17-18 November 2005 [DSTI/EAS/IND/SWP(2005)15]. This proposal can be seen as a pared-down version of option (IV) below, since the OECD is only involved in specifying the nature of the longitudinal data and storing the internationally comparable indicators.

A variant of this option exists if longitudinal surveys or one-off surveys with longitudinal questions are specifically carried out by NSOs under a coordinated research umbrella. The Eurostat project on 'Factors of Success' constitutes an example: an internationally harmonized business survey is put in place to assess the determinants of success for start ups. If the OECD Project on Entrepreneurship Indicators adopts a similar approach, it would also fall under this category.

Longitudinal in this context relates to data sets where individuals can be tracked over time, whether implicitly as snap shots or panels of datasets, with identification characteristics available, or explicitly, where the data set allows you to select individuals and view their responses/activity over time.

Option (II) Decentralised work with indicators based on 'snapshot' information

This option is similar to (I) in approach but less demanding in terms of national data as no longitudinal databases have to be constructed. The role of the OECD would be to propose and get agreement on indicators that can be constructed from business surveys or census data at snapshots in time.

Option (II) was also proposed at the Structural Business Statistics Expert meeting on 3-4 November 2005 and received a positive response from some countries. The construction of indicators of market structure by detailed industry that was proposed at the meeting would constitute a pilot study for the option. The proposal to link trade and business data at the enterprise level (discussed at the OECD Expert Meeting on Trade Statistics, 12-14 September 2005) is another example for this option, providing, for instance,

information on the import and export propensity (by partner country) of enterprises by size class. A small Steering Group will elaborate further details and a timetable will be drawn up.

Option (III) Special bilateral projects

This option includes the special case where OECD enters an agreement with one particular country to carry out micro-data analysis. A case in point is the study on productivity analysis for China put in place by the Economics Department: several officials from the National Bureau of Statistics spent several weeks at the OECD and analysed, jointly with OECD staff, micro data on Chinese firms from the economic census. In other words, micro-data resided at the OECD for a limited period of time and under the control of officials from the Chinese National Bureau of Statistics

Option (IV) Coordinated remote access

The coordinated access model corresponds to the suggestion made by Eric Bartelsmann (see STD/CSTAT(2004)12 and more recently DSTI/EAS/IND/SWP(2005)12). The model does not explicitly call for the actual transmission of data from NSOs to the OECD, rather it envisages the transfer of metadata and the creation of a Centre for micro-data that would allow the OECD Secretariat to write the programmes that will interface with micro-data at the NSOs. There are a number of ways in which this option could work in practice. One of them relates closely to option (V) below. The Centre would also need to be resourced in so far as a panel would need to assess the suitability of each resource proposal. A feasibility study will be carried out with a small set of volunteer countries.

Option (V) Transmission of micro-data to the OECD

Option (V:(1))

Examples for such transmission are national data that are provided to OECD (PISA) and Eurostat. There is a good model by the University of Minnesota (Integrated Public Use Micro-data Series, IPUMS) that compiles data on population censuses from various countries (44 countries have so far submitted census data). Extending these situations significantly will require legal agreements to be drawn up with countries relating to data transmission to the OECD and how it (and what) can be used. Given the volume of data and its confidential nature this option would need to fully explore the costs involved in setting-up and managing such a database; particularly concerning its use by OECD staff. One possibility would be to consider whether software such as μ and t-Argus, developed by the Dutch Central Bureau of Statistics and currently tested by Eurostat, could be used to preserve anonymity/confidentiality.

Option (V: (2))

The above examples are cases where national authorities transmit micro-data to the OECD. Another, possibly complementary, approach is to rely on private sources for micro-data. Commercial datasets exist for example for business data (such as AMADEUS) that originate from business information agencies such as Bureau van Dijk and Dun and Bradstreet. The biggest advantage of such datasets is that there are no issues of confidentiality. Disadvantages include the sometimes significant cost of purchase and unknown quality of the information. Commercial examples would be the publicly available private sources, such as Amadeus, Dun and Bradstreet etc.

SECTION 3 – OPTIONS (I) AND (II)

Option (I) Decentralised work with longitudinal data / Option (II) Decentralised work with indicators based on 'snapshot' information

Options (I) and (II) imply in many ways mechanisms that are little different to the data already provided to the OECD, for example the delivery mechanisms that govern the transmission of structural business statistics, which can be quite detailed, and, where in some cases, data cells are suppressed to preserve non-disclosure. In some respects therefore the approach is similar to the Data Cube approach already used in some countries. The disclosure issues that govern these two options are therefore somewhat different to those that prevail in options (III), (IV) and (V).

Indeed very little supplementary information is needed at this stage to develop these options further. It is clear for example that there are not likely to be issues of disclosure that prevent the transmission of indicators, in general, to the OECD, since, the ultimate objective of these options is to derive indicators that are disclosive. Granted depending on the degree of detail required it is possible that countries will not always be able to provide indicators. For example in some, particularly small, countries, information on high growth firms with specific characteristics (legal form, industrial activity etc) may constitute one firm only, and in these cases the data would normally be suppressed.

What is possible will become clearer as the feasibility study evolves and, in particular, when a list of target indicators is developed. In the meantime this section merely seeks to establish contact points and whether longitudinal data exists in 6 key data sets.

Question 3

Listed below are 6 of the most common sources of micro-data typically collected by national statistics institutes.

Micro-Data Description	Is the dataset longitudinal « Y/N »	If the dataset is not longitudinal – are there any plans to do so and if so when?	Contact details: e-mail
Population Census, or comparable administrative sources (PC)			
Labour Force Survey, (LFS)			
Statistical Business Register and/or administrative equivalents. (BR)			
Census/Surveys of businesses (manufacturing, construction, services etc) (SBS)			
External trade flows (ETF)			
Household Expenditure/Income Surveys (HES)			

SECTION 4 – OPTIONS (III, IV AND V)

Option (III) - Special bilateral projects / Option (IV) - Coordinated remote access / Option (V:(1)) - Transmission of micro-data to the OECD

This section focuses on the various forms in which micro-data can be made available to the OECD, and is of relevance to options (III) (IV) and (V:(1)). In this sense this section attempts to determine the form of micro-data, be it pure, anonymised, confidentialised, perturbed or artificial, (see Glossary for a description of terms) that can be made accessible to the OECD, depending on the access-mode, and the preconditions, (such as legal requirements, secure data transmission systems etc) that would need to be fulfilled beforehand by the OECD. It is important to note that for each of these options access to data is based on the notion that it will be restricted to the OECD Secretariat, who will be responsible for preserving confidentiality.

For convenience respondents have a choice of how they respond to the questions in this section. The first choice is to respond to a series of questions in Word. The alternative is to respond to the same questions in the attached Excel spreadsheet.

OPTION (III) - Special bilateral projects

4.(III)1 Population Census or comparable administrative sources

4.(III)1.1. Can data be provided to the OECD and in what form (pure set, anonymised, confidentialised, perturbed, artificial etc)? «Y/N, please specify the nature of the data set (level of confidentiality)»

4.(III)1.2 Under what circumstances can data be provided? Please describe the legal requirements/obligations that would need to be fulfilled by the OECD and any other preconditions, such as secure software systems

4.(III)2 Labour Force Survey

4.(III)2.1. Can data be provided to the OECD and in what form (pure set, anonymised, confidentialised, perturbed, artificial etc)? «Y/N, please specify the nature of the data set (level of confidentiality)»

4.(III)2.2 Under what circumstances can data be provided? Please describe the legal requirements/obligations that would need to be fulfilled by the OECD and any other preconditions, such as secure software systems

4.(III)3 Statistical Business Register and/or administrative equivalents

4.(III)3.1 Can data be provided to the OECD and in what form (pure set, anonymised, confidentialised, perturbed, artificial etc)? «Y/N, please specify the nature of the data set (level of confidentiality)»

4.(III)3.2 Under what circumstances can data be provided? Please describe the legal requirements/obligations that would need to be fulfilled by the OECD and any other preconditions, such as secure software systems

4.(III).4 Census/Surveys of businesses (manufacturing, construction, services etc)

4.(III)4.1. Can data be provided to the OECD and in what form (pure set, anonymised, confidentialised, perturbed, artificial etc)? «Y/N, please specify the nature of the data set (level of confidentiality)»

4.(III)4.2 Under what circumstances can data be provided? Please describe the legal requirements/obligations that would need to be fulfilled by the OECD and any other preconditions, such as secure software systems

4.(III).5. External trade flows

4.(III)5.1. Can data be provided to the OECD and in what form (pure set, anonymised, confidentialised, perturbed, artificial etc)? «Y/N, please specify the nature of the data set (level of confidentiality)»

4.(III)5.2 Under what circumstances can data be provided? Please describe the legal requirements/obligations that would need to be fulfilled by the OECD and any other preconditions, such as secure software systems

4.(III).6 Household Expenditure/Income Surveys

4.(III).6.1. Can data be provided to the OECD and in what form (pure set, anonymised, confidentialised, perturbed, artificial etc)? «Y/N, please specify the nature of the data set (level of confidentiality)»

4.(III).6.2 Under what circumstances can data be provided? Please describe the legal requirements/obligations that would need to be fulfilled by the OECD and any other preconditions, such as secure software systems

4.(III).7 Other datasets, which could be made accessible (please specify)

4.(III).7.1. Can data be provided to the OECD and in what form (pure set, anonymised, confidentialised, perturbed, artificial etc)? «Y/N, please specify the nature of the data set (level of confidentiality)»

4.(III).7.2 Under what circumstances can data be provided? Please describe the legal requirements/obligations that would need to be fulfilled by the OECD and any other preconditions, such as secure software systems

OPTION (IV) - Coordinated remote access

4.(IV).1 Population Census or comparable administrative sources

4.(IV).1.1. Can data be provided to the OECD and in what form (pure set, anonymised, confidentialised, perturbed, artificial etc)? «Y/N, please specify the nature of the data set (level of confidentiality)»

4.(IV).1.2 Under what circumstances can data be provided? Please describe the legal requirements/obligations that would need to be fulfilled by the OECD and any other preconditions, such as secure software systems

4.(IV).2 Labour Force Survey

4.(IV).2.1. Can data be provided to the OECD and in what form (pure set, anonymised, confidentialised, perturbed, artificial etc)? «Y/N, please specify the nature of the data set (level of confidentiality)»

4.(IV).2.2 Under what circumstances can data be provided? Please describe the legal requirements/obligations that would need to be fulfilled by the OECD and any other preconditions, such as secure software systems

4.(IV).3 Statistical Business Register and/or administrative equivalents

4.(IV).3.1 Can data be provided to the OECD and in what form (pure set, anonymised, confidentialised, perturbed, artificial etc)? «Y/N, please specify the nature of the data set (level of confidentiality)»

4.(IV).3.2 Under what circumstances can data be provided? Please describe the legal requirements/obligations that would need to be fulfilled by the OECD and any other preconditions, such as secure software systems

4.(IV).4 Census/Surveys of businesses (manufacturing, construction, services etc)

4.(IV).4.1. Can data be provided to the OECD and in what form (pure set, anonymised, confidentialised, perturbed, artificial etc)? «Y/N, please specify the nature of the data set (level of confidentiality)»

4.(IV).4.2 Under what circumstances can data be provided? Please describe the legal requirements/obligations that would need to be fulfilled by the OECD and any other preconditions, such as secure software systems

4.(IV).5. External trade flows

4.(IV).5.1. Can data be provided to the OECD and in what form (pure set, anonymised, confidentialised, perturbed, artificial etc)? «Y/N, please specify the nature of the data set (level of confidentiality)»

4.(IV).5.2 Under what circumstances can data be provided? Please describe the legal requirements/obligations that would need to be fulfilled by the OECD and any other preconditions, such as secure software systems

4.(IV).6 Household Expenditure/Income Surveys

4.(IV).6.1. Can data be provided to the OECD and in what form (pure set, anonymised, confidentialised, perturbed, artificial etc)? «Y/N, please specify the nature of the data set (level of confidentiality)»

4.(IV).6.2 Under what circumstances can data be provided? Please describe the legal requirements/obligations that would need to be fulfilled by the OECD and any other preconditions, such as secure software systems

4.(IV).7 Other datasets, which could be made accessible (please specify)

4.(IV).7.1. Can data be provided to the OECD and in what form (pure set, anonymised, confidentialised, perturbed, artificial etc)? «Y/N, please specify the nature of the data set (level of confidentiality)»

4.(IV).7.2 Under what circumstances can data be provided? Please describe the legal requirements/obligations that would need to be fulfilled by the OECD and any other preconditions, such as secure software systems

OPTION (V:(1)) - Transmission of micro-data sets to the OECD

4.(V).1 Population Census or comparable administrative sources

4.(V).1.1. Can data be provided to the OECD and in what form (pure set, anonymised, confidentialised, perturbed, artificial etc)? «Y/N, please specify the nature of the data set (level of confidentiality)»

4.(V).1.2 Under what circumstances can data be provided? Please describe the legal requirements/obligations that would need to be fulfilled by the OECD and any other preconditions, such as secure software systems

4.(V).2 Labour Force Survey

4.(V).2.1. Can data be provided to the OECD and in what form (pure set, anonymised, confidentialised, perturbed, artificial etc)? «Y/N, please specify the nature of the data set (level of confidentiality)»

4.(V).2.2 Under what circumstances can data be provided? Please describe the legal requirements/obligations that would need to be fulfilled by the OECD and any other preconditions, such as secure software systems

4.(V).3 Statistical Business Register and/or administrative equivalents

4.(V).3.1 Can data be provided to the OECD and in what form (pure set, anonymised, confidentialised, perturbed, artificial etc)? «Y/N, please specify the nature of the data set (level of confidentiality)»

4.(V).3.2 Under what circumstances can data be provided? Please describe the legal requirements/obligations that would need to be fulfilled by the OECD and any other preconditions, such as secure software systems

4.(V).4 Census/Surveys of businesses (manufacturing, construction, services etc)

4.(V).4.1. Can data be provided to the OECD and in what form (pure set, anonymised, confidentialised, perturbed, artificial etc)? «Y/N, please specify the nature of the data set (level of confidentiality)»

4.(V).4.2 Under what circumstances can data be provided? Please describe the legal requirements/obligations that would need to be fulfilled by the OECD and any other preconditions, such as secure software systems

4.(V).5. External trade flows

4.(V).5.1. Can data be provided to the OECD and in what form (pure set, anonymised, confidentialised, perturbed, artificial etc)? «Y/N, please specify the nature of the data set (level of confidentiality)»

4.(V).5.2 Under what circumstances can data be provided? Please describe the legal requirements/obligations that would need to be fulfilled by the OECD and any other preconditions, such as secure software systems

4.(V).6 Household Expenditure/Income Surveys

4.(V).6.1. Can data be provided to the OECD and in what form (pure set, anonymised, confidentialised, perturbed, artificial etc)? *«Y/N, please specify the nature of the data set (level of confidentiality)»*

4.(V).6.2 Under what circumstances can data be provided? Please describe the legal requirements/obligations that would need to be fulfilled by the OECD and any other preconditions, such as secure software systems

4.(V).7 Other datasets, which could be made accessible (please specify)

4.(V).7.1. Can data be provided to the OECD and in what form (pure set, anonymised, confidentialised, perturbed, artificial etc)? *«Y/N, please specify the nature of the data set (level of confidentiality)»*

4.(V).7.2 Under what circumstances can data be provided? Please describe the legal requirements/obligations that would need to be fulfilled by the OECD and any other preconditions, such as secure software systems

SECTION 5

Question 5.1

For all of the database-option combinations in section IV where the pure set could not be made available please describe below why your institution would not be able to provide this data, citing costs, resources, legal restrictions etc as may be the case, as reasons. Please also describe the reasons for not being able to provide anonymised or confidentialised data as may be the case.

Option	Micro-Data Set
	<i>Description of Micro-Data Set 1</i>
III	<i>Reason</i>
IV	<i>Reason</i>
V(1)	<i>Reason</i>
<i>Description of Micro-Data Set 2</i>	
III	<i>Reason</i>
IV	<i>Reason</i>
V(1)	<i>Reason</i>
<i>Description of Micro-Data Set n</i>	
III	<i>Reason</i>
IV	<i>Reason</i>
V(1)	<i>Reason</i>

Question 5.2

Please rank options (I) to (V:(2)) in order of preference (ranking each option from 1-6: with '1' representing the most preferable option and '6' the least preferable).

Option	Micro-Data Set						
	<i>PC</i>	<i>LFS</i>	<i>BR</i>	<i>SBS</i>	<i>ETF</i>	<i>HES</i>	<i>Other</i>
I							
II							
III							
IV							
V:(1)							
V:(2)							

ANNEX B: MICRODATA ACCESS BY INTERNATIONAL ORGANISATIONS/AGENCIES

CES TASK FORCE ON CONFIDENTIALITY AND MICRODATA ACCESS

Summary

The discussion at the Conference of European Statisticians in June 2005 concluded that more information was required on the provision of microdata to international organisations and researchers from other countries to better prepare the Guidelines⁷. To meet this objective, a survey questionnaire detailing the terms and conditions of microdata access offered by countries to international organisations and researchers from other countries was sent to 61 countries. Forty three countries responded to the survey. Presented below is a summary of findings.

PART A

Question 1: Microdata Release to International Organisations

Of the 43 countries, only three countries do not release anonymised microdata to international organisations. These are Japan, Ukraine and the Republic of Moldova.

Of the 40 countries that release microdata to international organisations, 25 countries (62%) place no limitation on the organisations. The United States and Canada do not place any limitations on organisations for anonymised public use files. Sweden does not draw a distinction between international organisations and research institutions. In exceptional circumstances, Australia is also able to release microdata files to their counterpart agency of a foreign government, provided the rules of confidentiality are met. Ireland places no limitations on organisations for anonymised microdata from household surveys. Germany and Greece provide microdata on Community Statistics to Eurostat under the provisions of the Commission's Regulation. Andorra did not respond to whether it releases data to international organisations.

Fourteen countries (35%) place limitations on organisations of which five release data to Eurostat only. These are Austria, Denmark, Norway(?), Portugal and Russia (?). Other countries such as Hungary and Macedonia release data to international organisations engaged in statistical activity, Lithuania to Eurostat and scientific research organisations and Romania to Eurostat and research institutes.

OECD: Of the 40 countries that release data to international organisations, 23 countries (57%) belong to the OECD. Of these 23 countries, 16 place no limitations on the organisations to which they release microdata.

Question 2: Forms of Microdata

In the survey, three main methods of releasing microdata were suggested and countries were asked to fill in the forms in which they *could* release microdata. Of the 40 countries that release data, 55% (22

⁷ Managing Statistical Confidentiality and Microdata Access-Principles and Guidelines of Good Practice.

countries) release anonymised microdata files as public use files, 82% (33 countries) release anonymised microdata files as licensed files and only 12% (5 countries) have remote access facilities. These are Australia, Canada, Germany, Ireland and United States. In Canada only non-confidential aggregate data are released via remote access facilities. New Zealand mentioned that it was currently developing its remote access facilities. Fifteen countries are able to release anonymised microdata as both public use and licensed files, while two countries, Ireland and United States (US), can release data in all three forms. However, the US is only able to release licensed files and offer remote access facilities within their borders.

OECD: Amongst the 23 OECD countries that release data to international organisations, 14 are able to release microdata as public use files, 19 as licensed files and 5 have remote access facilities.

Question 3: Forms of Microdata Actually Released

Following the question on available forms of microdata, countries were asked the forms in which they *actually* released microdata to international organisations. Of the 40 countries that are able to release data to international organisations, 16 countries (41%) actually released data as public use files. Licensed files were the most popular mode of actually released data with 27 countries (67%) responding affirmatively. Three countries release data via remote access facilities. These are Australia, Canada and Germany. In Canada only non-confidential aggregate data are released via remote access facilities.

OECD: Of the 23 countries, 12 countries actually release data to international organisations as public use files, 15 as licensed files and 3 with remote access facilities.

Question 4: Specific Conditions

Countries were surveyed on the conditions that applied to the provision of microdata. Of the 40 countries that responded affirmatively to releasing data to international organisations, 18 countries (45%) place no *particular* conditions while 20 countries (50%) apply conditions on the release of microdata to international organisations.

Armenia, Australia, Bulgaria, and Hungary state that they can release data only for statistical purposes. Austria, Denmark, Greece and Italy specified that the data can be released as per the Council Regulations 831/2002. Estonia and Mongolia require a contract for the transmission of data for the purposes of scientific research and confidentiality respectively. Ireland mentioned that although in principle it could release anonymised data to international organisations, in practice it would only release anonymised microdata from household surveys. Israel specified that the standard PUF (public use files) agreement must be signed by international organisations. San Marino and Sweden can only release aggregated data to international organisations. Switzerland stated that the data could not be transmitted outside the international organisation using it and must either be returned or destroyed after use. Macedonia requires that the data released should be in accordance with their law on data protection under the purview of the State Statistical Office. United Kingdom mentioned that it could either provide fully anonymised data to countries outside Europe or data under contract. For countries within Europe, data could be provided under contract with other EU states or under condition R381. United States specified that to obtain access to data from the Bureau of Labour Statistics (BLS), individuals from international organisations would have to become a BLS agent, which was only possible if the individual had a social security or alien registration number.

Andorra and Norway did not respond to this question. Although Ukraine does not release data to international organisations, it specified that the provision of microdata to international organisations would have to follow the Law of Ukraine on State Statistics, Article 21, 22.

OECD: The group of OECD countries was evenly divided over the application of specific conditions, with 11 countries imposing no particular conditions and 11 countries applying specific conditions. Norway did not respond to the question.

Question 5: Data Release: Organisation, Division, Individual

Countries were requested to specify whether they were able to release data to the whole organisation, a specific division or a designated individual. Of the 40 countries that release data to international organisations, 21 (52%) can release data to a whole organisation, 20 (50%) to a division and 21 (52%) to a designated individual. Seven countries (17%) can release data to all three audiences. These are Armenia, France, Germany, Kazakhstan, Mongolia, Slovenia and United States. The answer provided by Ireland was unclear while San Marino and Switzerland did not respond to this question.

OECD: Of the 23 countries, 14 countries can release microdata to a whole organisation, 9 countries to a particular division and 11 countries to an individual.

Question 6: Business Data Arrangements

Countries were surveyed on whether they adopted different practices when releasing data on businesses. Of the 40 countries able to release data to international organisations, 26 countries (65%) remarked that they do not follow any different arrangements for business data. Thirteen countries (32%) reported on following different procedures. Andorra reported on having no data on businesses.

Canada does not produce microdata files for businesses. Latvia only releases data relating to persons. New Zealand is able to release business data in principle but does not in practice due to the difficulties in ensuring confidentiality. Moldova, which falls outside of the 40 countries that release data to international organisations, reported on not having any different arrangements for business data.

OECD: Of the 23 countries, 7 countries indicated that they followed different arrangements for business data while 16 reported on no such differences.

Question 7: Data Release to Citizen Researchers collaborating with an International Organisation

In the questionnaire, countries were asked to spell out if they were able to release microdata to researchers from their own country working in collaboration with an international organisation. Of the 40 countries, 34 countries (85%) responded positively. This includes Australia, which has recently amended its legislation to permit data recipients to collaborate with other individuals and organisations. Canada releases public use files available to the general public. In addition, subject to certain conditions, Canada allows its researchers to gain access to confidential microdata available in its Research Data Centres. Germany is able to release microdata as long as the data sets were physically located in Germany with no possibility to cite the microdata from abroad. In this case scientific use files that correspond to licensed files can be released. United Kingdom can only release public use files. The United States specified that it permits the researcher to share the output from the data, not the data itself.

Six countries (15%) are not able to release data. This includes Lithuania which can only release data to research institutions.

Although Ukraine and Moldova do not release microdata to international organisations, they do release data to researchers in a manner that protects identification of data.

OECD: Of the 23 countries in the OECD that release data to international organisations, 20 are able to release data to their citizen researchers collaborating with international organisations. Three countries are not able to release data.

Question 8: Comments

Countries were requested to provide additional comments to the questions asked.

The **Czech Republic** sends statistical information abroad, free of charge, only when sent on a mutual basis or when stipulated by an international agreement or convention. For other cases, the Republic's Statistical Act regulates payment for statistical information sent abroad.

France sends confidential business data outside European Regulation to Eurostat only.

To ensure that the data released is not used for purposes outside of those agreed, the Division of Statistical Information and Editions in **Greece** takes the necessary administrative, technical and organisational measures. Greece is also able to provide on-site facilities to national and international researchers.

Hungarian legislation does not use the term microdata, instead it employs the terms individual and statistical data. Individual data pertain to data that serve a statistical purpose and which relate to a natural or legal person or to a data supplier without legal personality. This data can only be released to international organisations that are engaged in statistical activity. Furthermore, aggregated data cannot be made public where there are data about less than three respondents in a cell. The exception to this rule is the release of data to international organisations engaged in statistical work, which requires prior written consent of data providers where they number less than three.

Israel is able to release public use files to anyone, which includes international organisations. Microdata files under contract (MUF) are released only to specified categories of recipients in Israel.

In **New Zealand**, published aggregated data are supplied to international organisations by Statistics New Zealand. For the supply of microdata to international organisations, it is required that the provisions of the microdata access protocols described in the New Zealand Statistics Act (1975) are followed. These apply not only to international organisations but to all users requesting access to microdata.

In **Poland**, for individual cases, the President of its Central Statistical Organisation supported by the Statistical Confidentiality Commission, can take decisions regarding the dissemination of microdata.

Romania reiterates that anonymised microdata is released only for research purposes with special attention to the fulfillment of the rules of confidentiality under "disclosure control."

In **Russia**, microdata are transformed when releasing it to international organisations to prevent identification.

Serbia and Montenegro qualify their responses filled in the survey by stating that it reflected their position on microdata release despite few such requests from international organisations.

Japan, a country that does not release microdata to international organisations, stated that it was currently studying the system for producing and disseminating microdata in terms of legal and technical aspects.

Although the **Republic of Moldova** does not release data to international organisations, it stated that access to confidential information that did not jeopardise confidentiality could be granted by the director general of the central statistical body for research projects and programs of national importance. Furthermore, confidential information for scientific purposes was granted to persons directly engaged in the project, provided that confidentiality was not breached.

PART B

Question 9: Arrangements for Statistical Officers from other Countries

Countries were asked whether there were any different arrangements possible for statistical officers from other countries. Of the 40 countries that provide microdata to international organisations, 32 (80%) follow no different arrangements. This includes the countries of France and Sweden, who annotated their responses. France commented that its “comité du secret” often employs a “reciprocity clause” to ensure that French researchers/statisticians can access data from the other recipient countries. In Sweden, in some cases it was possible to exchange microdata licensed files concerning community statistics with other statistical agencies in the European Union. Andorra and San Marino did not respond to this question.

Six countries (15%) cited different arrangements. These were Canada, Czech Republic, Norway, Poland, Serbia and Montenegro and United Kingdom.

San Marino and Andorra did not respond to the question.

OECD: Eighteen of the 23 countries indicated that they had no different arrangements possible for statistical officers from other countries. The remaining five countries pursued alternate arrangements.

Question 10: Microdata Release to Researchers in Other Countries

In the survey, countries were asked whether they could release microdata to researchers in other countries. Further, if they were able then in which forms was this data released. Of the 40 countries that release microdata to international organisations, 31 countries (77%) replied positively whereas 7 (17%) replied negatively. Some countries such as Hungary and Ireland, do not release data directly but through agencies. Canada can only release public use files. Lithuania can release microdata to research institutions only, whereas United States can only do so for researchers within its territory. The Czech Republic requires a bilateral agreement. In Poland, the President of the Central Statistical Organisations can take decisions regarding microdata release on a case by case basis. Each researcher must provide scientific documentation approved by the concerned institute. In Norway, only licensed anonymised data can be supplied. Sweden requires that conditions of confidentiality be met and that the data user provides a description of how confidentiality would be secured. In addition, Sweden imposes limitations on access to microdata by specifying beforehand the people, projects, variables and periods during which the data can be used.

Andorra and San Marino did not respond to the question

On the forms of data released, inferences were drawn for those countries that did not reply. These inferences are based on the questions asked earlier on forms of microdata available and actually released to international organisations. Of the 31 countries that replied positively to releasing data to researchers abroad, 14 are able to release microdata as public use files (16 countries replied negatively), 21 as licensed

files (9 replied negatively) and 3 had remote access facilities (27 replied negatively). Furthermore, Denmark, Italy, New Zealand, Poland and United States had on-site facilities.

OECD: Eighteen of the 23 countries of the OECD group indicated that they could release data to researchers in other countries. The remaining five countries could not.