For Official Use **STD/CSTAT/RD(2006)2**

Organisation de Coopération et de Développement Economiques
Organisation for Economic Co-operation and Development **19-May-2006**

English - Or. English

STATISTICS DIRECTORATE
COMMITTEE ON STATISTICS

STD/CSTAT/RD(2006)2
For Official Use

OECD CONFERENCE ON, AND FEASIBILITY STUDY OF, MICRODATA

to be held on 12-13 June 2006
at the Espace Louis Armand
UICP, 16 rue Jean Rey, Paris

The development of macro-based statistics and indicators has rightly formed the focus of statistical development and resources, particularly in an international context, for many decades now but as the world becomes increasingly complex and computational capabilities increasingly advanced it is also right to start considering whether the role of micro-data in international statistics should not also be increased. This paper describes the OECD's recently started project to assess the feasibility of accessing and using national and international data sets of micro-data or indicators derived from them. Five options will be investigated in this study, varying in degrees of complexity both from a technological, legal and resource perspective. Typically however the greater the degree of complexity the greater the potential benefits and applications. Also announced in this paper are OECD plans for a conference on micro-data, to be held in November of this year, that is intended as a forum to share experiences of micro-data use and access and that is hoped to kick-start a process that will lead to more systematic development of international micro-data sets and/or new internationally harmonised indicators based on, often readily available, micro-data.

The Committee is invited to express its views on these developments.

For more information, please contact:
Nadim Ahmad, Tel: +33.1.4524.8847 or Email: nadim.ahmad@oecd.org

JT03209238

English - Or. English

OECD CONFERENCE ON, AND FEASIBILITY STUDY OF, MICRODATA

Background

1. Indicators and analyses based on micro-data constitute a promising avenue for new research and value can be added by deriving internationally comparable results. There is experience at the OECD with coordinating research projects based on micro-data but such efforts have generally been on a one-off basis without consideration to establishing a more permanent framework. A first proposal for such a framework for firm-level data was presented to 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 and to organise a conference on micro-data (uses and access) where the results of this study could be presented and experiences of national and international institutions could be shared.

2. Len Cook, formerly the Director of the United Kingdom's Office for National Statistics, has been appointed as a consultant to take forward the feasibility study and to produce a report describing the outcomes of this study. This present note outlines a structure for the report of this study, along with some ideas on how the work could fit in with ongoing projects in business statistics and productivity analysis, together with more information on the nature, timing and location of the conference. It is important to state up-front that the study and conference will not limit themselves to OECD countries and, so, many, particularly large, non OECD countries, such as Brazil, China, India, Indonesia and South Africa, will also be included in the scope of this study and the conference.

Objectives of the feasibility study report

3. The report will outline and assess the various options for OECD to advance on micro-data work at the international level. It should provide a basis for the Statistics Committee to make an informed decision about follow-up activity.

Basic structure of the report

Chapter 1: Background

4. This chapter sets the stage. It would discuss the importance of micro-data analysis and how it complements more traditional work with aggregate data. Reference could be made to international studies on firm and non-firm level micro-data carried out by OECD and other international organizations, for example the World Bank. Generally, the background chapter should define and delineate the scope of micro-data work and present information on the needs of users and the OECD.

5. The background chapter should also make reference to other international activities in the area of micro-data such as CAED (academically-oriented analysis), work on data confidentiality by the UN-ECE (IT and producer-oriented issues of micro-data access).

6. Reference should also be made to current practice, policies and examples of micro-level access in national statistical offices and to recommendations on access to micro-data for research that are in preparation at the CES.

Chapter 2: Review of selected international practices

7. Chapter 2 would take a look at some cases where micro-data has been made available and/or used for analysis at the international level. The chapter takes stock of existing practices, for example:

- Eurostat
- IPUMS
- Luxembourg Income Study
- PISA
- Inventory of confidential data transmitted to the OECD

Chapter 3: Main options to carry out firm-level micro-data work at the OECD

8. This chapter assesses the main options for micro-data work at the OECD. For every option, the following questions will be asked. (i) What would be the analytical usage? (ii) What is the legal feasibility? (iii) What are the costs and benefits for OECD and for national statistical offices?

Option (a) Transmission of micro-level data sets to the OECD

9. Examples for such transmission are national data that are provided to OECD (PISA) and Eurostat or the 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). This option 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 and, possibly, outside researchers, where, certainly for the latter group, contracts would need to be drawn-up, including scope for enforceable penalties for deliberate or accidental disclosure of confidential information. 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.

10. The above examples are cases where national authorities transmit micro-data to the OECD. An alternative option is to rely on private sources for micro-data. Commercial datasets exist for example for business data (such as AMADEUS) that originate from credit rating agencies such as 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.

Option (b) Coordinated remote access

11. 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 researchers 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 (a) above. All options however would require that contracts would need to be drawn up between the research centre, on behalf of NSOs, and the researchers.

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 two or three volunteer countries. Giuseppe Bruno from the Italian Central Bank will work with the OECD's Information Technology and Network Services Directorate for two months, beginning in June, to assess the feasibility of this option.

Option (c) Decentralised work with longitudinal data

12. This is the option that has been followed by OECD in previous longitudinal studies (see work by OECD's Science Technology and Industry and Economics Directorates); 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 were discussed by the Statistical Working Party of the Committee on Industry and Business Environment (SWIC) on 17-18 November 2005 [DSTI/EAS/IND/SWP(2005)15]. These proposals can be seen as a pared-down version of option (b) above, since the OECD is only involved in specifying the nature of the longitudinal data and storing the internationally comparable indicators.

13. 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.

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

14. This option is similar to (c) 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.

15. Option (d) was also proposed at the OECD's Structural Business Statistics Expert meeting on 3-4 November 2005 and received a positive response from at least a few 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 (OECD's Expert meeting on Trade Statistics, 12-14 September 2005) is another example of 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 (e) Special bilateral projects

16. This option constitutes a 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 (NBS) 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 NBS.

Chapter 4: A set of rules and procedures for OECD

17. Somewhat independent from the outcome of the assessment of the various options in Chapter 3, a set of rules and procedures should be elaborated for the management of micro-data at the OECD, if such data are transmitted to the OECD.

18. Such rules should be connected to the existing OECD quality framework and possible synergies should be explored with the rules and procedures that are in place at Eurostat.

19. Finally, this chapter could also outline some IT solutions for accessing and transmitting micro-level data.

Process and timing

20. A steering group of internal OECD experts comprising those with experience of acquiring micro-level data and users of micro-level data has been established. The immediate next steps are the development and circulation of a questionnaire that will attempt to ascertain in a relatively quick way the feasibility of each of the options listed above and the preference for each of the options of national statistics institutions. As such its primary aims are: to establish the availability of micro-level data in statistics institutes; the legal constraints imposed on the transmission and use of micro-level data outside of national institutions; and the current mechanisms, if any, used to disseminate micro-level data including the applications (software) and methods, such as anonymisation processes, used in dissemination. It is planned that this questionnaire will be circulated to national statistics institutes in the beginning of June.

21. The returns to the questionnaire will form a large part of the final report, a draft of which will serve as an input into the planned micro-data conference in November 2006. The final report will be prepared shortly after the conference.

Conference on micro-data

22. The catalyst for the conference is the feasibility study described above. But the conference is intended to be more than just a vehicle to present the findings of this study. Indeed it is intended that the conference will be an opportunity to share experiences across institutions of the benefits and problems inherent in micro-data analysis and access. Statistical development over the last few decades has rightly focused on macro-based statistics but throughout this period, and increasingly so in more recent years, the role and benefits of micro-data in assisting policy analysis have grown and become more visible, particularly in the context of globalisation and studies of productivity, growth and entrepreneurship, for example. Coincident with this is the vast improvement the world over in statistical information systems and the time is ripe to capitalise on these developments and begin to think more systematically outside of the ‘macro’ box.

23. Clearly some of the options investigated within the feasibility study are more challenging than others, particularly option (a) but it should be possible at the very least, to tap into micro-level data by developing indicators that are commonly used but not currently internationally harmonised, without compromising confidentiality, such as envisaged in options (c) and (d). Indicators such as GDP, for example, have long been subject to internationally comparable rules but other indicators that can help policy makers understand and improve economic performance such as the percentage and characteristics of high growth firms; indices of firm concentration; economic activity broken down by ownership type; business exits, business entries, to name but a few, although commonly available are not, typically, harmonised internationally although the literature contains many examples of their use. Moreover much of this information, for example, can be sourced from statistical or administrative business registers which all OECD countries have, and so the scope to develop many indicators globally is evident.

24. An exact date or venue for the conference is yet to be finalised. The conference will however be held in Europe for two, possibly three, days in November. The exact timing will be shortly after or before the related OECD meetings of the Statistical Working Party of the Committee on Industry and Business

Environment (SWIC) (13-15 November) and Structural Business Statistics Experts (9-10 November) to maximise the participation of delegates from outside Europe.

25. A draft outline of subject areas that the micro-data conference will tackle is shown below.

- **Session 1 -**

  Applications of micro-data – national and international examples of micro-data analysis; benefits; what can micro-data reveal that macro-data cannot and do the messages differ; longitudinal micro-data sets and snap-shots, pros and cons; commercial micro-data sets – experiences, pros and cons; are these better suited to dealing with globalisation issues such as the operations of multi-nationals.

- **Session 2 –**

  Micro-data access in practice – examples of micro-data access in a national and international context:- remote access, safe centres, one-off applications, survey approaches; data cleaning, synthetic datasets, and preserving anonymity.

- **Session 3 -**

  The OECD Questionnaire on micro-data – synthesis – conclusions, review of country practices; data-content, preferences. Providing an inventory, and comparing sources, of micro-data – administrative data (registers), surveys, censuses – scope for harmonising indicators.

- **Session 4 -**

  The OECD feasibility study - A review of the five options: how feasible are they and in which countries; an assessment of the feasibility of option (a) focusing on legal and technological constraints; results of the pilot study of option (b), co-ordinated remote access; standard indicator development (options (c) and (d))

- **Session 5 -**

  Moving forward and developing an action plan .

26. A more detailed agenda will be developed over the coming weeks and will be publicised more widely at the earliest convenience. Those interested in presenting at the conference are invited to send expressions of interest, together with an outline of their proposed presentations to nadim.ahmad@oecd.org.