STATISTICS DIRECTORATE

SBS Expert Meeting "Towards better Structural Business and SME Statistics"
OECD, Statistics Directorate
3-4 November 2005
La Muette Room 4

Agenda Item 5 : Facilitating data delivery for OECD's SBS programme

Mr. N. Ahmad, OECD

REDUCING THE DATA-DELIVERY BURDEN

Exploring Options to Improve the Quality of Data Dissemination

Nadim Ahmad, Statistics Directorate

OECD Structural Business Statistics Expert Meeting – 3-4 November 2005

Overview

1. In recent years the OECD Statistics Directorate has increased its coverage of structural business indicators to include more comprehensive coverage of activity, in particular the service sector, to include indicators providing information relating to activity by size class, and, to expand the set of economic variables collected. These developments, in response to growing policy needs such as entrepreneurship policy[1], continue at pace, and the OECD Statistics Directorate is currently assessing the feasibility of developing further complementary indicators in this field, including the creation of a database on business demography statistics.

2. Clearly these increasing demands imply an increased burden on statistical offices. Naturally, statistical offices collect business information based on the statistical institutional arrangements in place in their country but this can often present difficulties when international comparisons are required. In the case of business statistics there are many reasons why this may occur - for example some countries are only able to provide data for some sub-sectors of the economy, such as manufacturing say, or the types of businesses covered differ depending on size thresholds that may be present in business and administrative registers. Fortunately metadata describing these differences is nearly always available and so users are able to correctly compare like with like.

3. One area however where difficulties can and do occur concerns the different industrial classification systems used across countries. OECD structural business statistics databases and many other OECD databases (for example the national accounts database and the structural analysis database) are based on the International System of Industrial Classification Revision 3 (ISIC Rev 3). The means by which countries (and the OECD) convert national data using national classification systems to ISIC Rev 3 is based on concordance tables describing the relationships between the two systems.

4. In many cases however, especially at the 4 and lower digit classification levels, the relationships between the different activities are one-to-many[2]. In other words, one activity in a national classification system corresponds to a number of activities in the ISIC system. This means methods are needed by which national data can be converted from single activities to multiple ISIC activities.

---

[1] The Bologna Charter on SME Polices and the 'Bologna Process'; The 2nd OECD Ministerial Conference on SMEs, Istanbul 2004; European Commission Lisbon Summit and the "Lisbon Strategy', 2000.

[2] Many-to-one relationships are also common although these present fewer difficulties that one-to-many relationships when the ultimate aim is an ISIC based information set.

5.   There are many ways in which these conversions are done in practice. The simplest approach is to simply allocate all firms[3] in a specific national classification group to its most likely ISIC classification group based, usually, on the ISIC activity from which most turnover is likely to come, in other words by creating one-to-one (and many-to-one) relationships between the classification systems. More sophisticated approaches also exist. For example some national agencies carry out the same type of allocation but with supplementary information on size-class and/or legal form, such that firms within one activity group in a national classification system are allocated to various ISIC groups in a systematic way based on their characteristics.

6.   Both approaches present compilation difficulties where international comparisons are concerned. Users want consistency across all data sets that refer to business statistics; meaning that all business datasets for a specific country should use the same classification conversion techniques. Even for the most simple of cases it is not certain that this happens in practice. The OECD receives[4] information by ISIC industry for a number of variables, size-class, foreign affiliates, value-added, turnover, employment, and this information is often based on different data sources and provided by different national agencies. Without explicit and common rules (within each country) on how the concordances are drawn up therefore the risks of inconsistent approaches across datasets for the same country are high.

7.   Moreover, even where consistency in approaches can be assured, significant care will still be needed when allocating firms on the basis of characteristics, for example size or legal form. A great deal of entrepreneurship policy, for example, relates to these very two characteristics. The risk therefore is that statistics based on preconceptions relating to the types of firms that operate in different industries, merely reinforce these preconceptions rather than reflect the actuality.

8.   But the difficulties do not stop here. All of this presupposes that the conversion routines are simply a case of determining the relationships and applying the rules that govern these relationships but confidentiality rules severely complicate matters. In providing ISIC data to the OECD, countries need to inure that non-disclosive cells cannot be deduced if data presented with a country's own-classification system and ISIC are combined. It's important to note that this statement primarily relates to non-EU OECD countries. For EU countries, Eurostat and the OECD have a Memorandum of Understanding that governs the transmission of data for EU countries. The approach adopted by Eurostat is to take 4-digit NACE information provided by countries and use concordances at this level to provide the OECD with ISIC information. This does mean, in practice, that data is, on some occasions, over-confidentialised. A fact confirmed by the loss of some 4-digit data in the OECD's database since the MoU came into effect. Ideally the concordance would be done at a lower level, NACE 6-digit to ISIC-4 digit for example but this is not as easy as it sounds.

9.   For Eurostat and non-EU OECD countries insuring that non-disclosive cells cannot be deduced by combining national data presented with a country's own-classification system and ISIC information can be time-consuming, since it amounts to a second confidentiality exercise. Given these burdens, and the (often severe) penalties in some countries for unintentional disclosure of non-confidential cells, it is perhaps understandable that data presented to the OECD often undergoes a higher level of confidentialisation than is theoretically necessary, resulting in reduced functionality and cross-country comparability of OECD databases and often a lack of timeliness in data delivery to the OECD.

10. The aim of this paper is to propose that the OECD intensifies bilateral discussions with countries in order to consider approaches, together, that minimise this second level of confidentialisation, improve

---

[3] Where firms are defined on the basis of the statistical unit in place, for example enterprise, establishment, legal unit etc.

[4] And converts national data based on national classification systems

cross-country comparability and comparability across databases that contain structural business statistics based on ISIC, whilst at the same time reducing the burden of national agencies. For some countries, of course, the net result of this consultation exercise will imply no change to the current transmission programme.

Present Situation

11. Currently the OECD structural business statistics databases (SEC and SSIS) provide information on a number of economic variables such as value-added, turnover, production for the business sector (largely) by 4-digit ISIC category and by enterprise/establishment size class. Additionally a number of other OECD databases, such as the national accounts database and the STAN database contain similar information, but usually by 2 to 3 digit sectors.

12. The quality and coverage of data delivered to the OECD's two SBS databases varies widely by country and by dataset. Moreover, for some countries, there are differences between the two datasets even where the concepts are the same. This partly reflects differences in timing[5] but the differences are so large on some occasions that other factors must also play a role. Finding out why these differences exist is obviously important. The paper itself will not present these differences as they are too numerous to mention but the OECD Statistics Directorate will cooperate bilaterally with countries to see what measures could be take to improve the coherence of these datasets.

13. Additionally the coverage of 4-digit sectors by country in the OECD's structural business statistics databases varies widely. Sometimes this reflects the lack of any data at the 4-digit level (in other words no concordance relationship between a country's classification system and ISIC for this 4-digit industry), on other occasions however it reflects over-confidentialisation. There is little the OECD can do about the former. However, exploring what can be done to improve the latter is worthwhile.

14. The idea is that the OECD will work with statistical agencies responsible for providing data to the OECD and explore with these offices whether the data transmission programme for SBS could be improved if agencies delivered data in their most detailed form on the basis of their own classification system.

15. For those countries which struggle to meet OECD requirements and that over-confidentialise data, one approach worth considering is whether the OECD should convert national data into ISIC Rev 3 (and in time ISIC Rev 4) using concordance relationships that are consistent with business statistics data used in other economic databases, particularly the OECD national accounts database. Inevitably some of the national data supplied to the OECD will be confidentialised, so these countries would additionally need to provide SBS data in its entirety for all digit levels used in their own national classification system (2-digit, 3-digit, 4-digit 5-digit etc). In this way the OECD statistics Directorate, in collaboration with the national statistical office, will be able to determine the most appropriate concordance relationship that minimises the amount of confidentialised cells, whilst at the same time maximising the amount of 4 digit information available.

Concordance Relationships

16. Two commonly used approaches to convert one classification system into another were briefly described above. The first and simplest way amounts to creating one-to-one and many-to-one relationships

---

[5] The Questionnaires are currently sent out at different times, although there are plans to change this for next year's data submission.

between a country's classification system and ISIC Rev 3. The second is a more sophisticated approach that allocates firms to an ISIC class on the basis of additional criteria, such as legal form and size class.

17. Depending on what is done in producing other business based statistics, such as gross value-added by industry for the national accounts, the OECD Structural Business Statistics Team's preference is for the first approach because it is simpler to implement and makes no assumptions about the types of firm (size, legal class) that are to be found in different activities.

18. Arguably this may not be the best approach. There is a risk that the comparability of datasets across countries will be compromised where there are several one-to-many relationships that need to be converted to one-to-one relationships. This may mean, for example, that in one country all firms classified to an industry group producing jewellery including watches are allocated to ISIC 3691 (*manufacture of jewellery and related articles*), whereas in another the same group is allocated to ISIC 3330 (*manufacture of watches and clocks*). If the concordance conversions are done within the OECD however, or at least better understood by the OECD, it will at least be possible to state exactly where these issues of (strict) incomparability arise.

19. For some countries, this approach will not be necessary and indeed may lead to a reduction in the level of detail currently available on the OECD's structural business statistics databases, as these countries are able to efficiently build up ISIC data from the firm level, meaning that the number of disclosive cells is kept to a minimum. This means that the approach to conversion will need to be developed on a country-by-country basis. Where the transmission of data works smoothly the OECD does not propose any changes to the transmission program, however the process of consultation with the OECD will at least mean that the OECD SBS team will better understand the concordance relationships used by these countries.

Benefits

20. The benefits of adopting this more flexible approach to data transmission are many. Clearly for those countries that struggle with data transmission the direct benefit is a reduction in the data burden. For the OECD the benefits are: more timely data; greater internal consistency between SSIS and SEC and possibly the national accounts and STAN; potentially more detailed, less confidentalised, data; and the ability to produce data that is more consistent over time as national and international classification systems change.

21. This last benefit is of considerable importance since a new international standard for industrial classifications ISIC Revision 4 is planned in 2007; indeed a draft version is already available. Typically when these changes occur it is difficult for the OECD to provide consistent time series as countries are generally reluctant to resubmit historical data on the basis of new classification systems, and the changes between the old and new systems are often too large for the OECD to convert old series to new on the basis of data submitted at the ISIC level. Of course it is not impossible to link the old and new series but there is a risk that, in doing so, inconsistencies remain. The risks are likely to be smaller however, or at least more manageable, if the OECD is responsible for transforming national data into ISIC.

Case Study – United States

22. To illustrate how this approach could improve the situation for some countries. The OECD Statistics Directorate recently completed a feasibility study that took size class information available from the US Census Bureau's website and transformed this data from NAICS[6] into ISIC Revision 3. The approach and results are presented below.

---

[6] North American Industrial Classification System

Background to US data

23. At present the OECD SEC and SSIS databases are not entirely consistent for US data, although the US is not exceptional in this regard. Generally, the availability of data by 4-digit industry before 2001[7] varies considerably between the two datasets as does the coherence of data in one case (wages and salaries of employees, where some differences exist). For example, the SEC dataset for 2001 data contains information at the 4-digit level for 21 industries only, and at the 3-digit level, information is available for only 20 industries. This compares with well over 100 for both the 3 and 4 digit information in the SSIS database.

Concordance

24. Part of the reason for the reduction in the 4-digit ISIC coverage of the SEC data compared to the SSIS data reflects the fact that the number of disclosive cells in the 6 digit US NAICS dataset is greater; not surprisingly given the greater level of detail afforded by the size class breakdown. Consequently the conversion from 6 digit NAICS to 4-digit ISIC is complicated, especially since it is necessary to ensure that unintentional disclosure does not occur.

25. In order to investigate what role the OECD could take in assisting countries in compiling their statistics the OECD Statistics Directorate recently completed an exercise that attempted to build up one-to-one (and one-to-many) concordance relationships from the 4-digit NAICS categories to the 4-digit ISIC categories. From the outset it was known that making the concordance at this level would inevitably result in some inconsistencies between the SSIS dataset (which is built up from the 6-digit NAICS level, and which suffers less from confidentiality problems) and the SEC dataset[8] but the nature of the exercise was experimental and to examine various approaches.

26. The ultimate objective of the exercise was to see if more detailed 4-digit SEC information could be provided by building up concordances from the 4-digit NAICS level. However a secondary objective was to assess the impact of improving the coherence between the SEC database and the OECD's national accounts and STAN databases, where concordance relationships between NAICS and ISIC are also developed at the 4-digit and above NAICS level.

27. The results of the exercise have been promising, resulting in significantly more 4-digit cells than is currently the case for pre 2001 US data on the SEC database. The cost, of course, is inconsistencies with a considerable number of 4-digit industries in the SSIS database – although this can be easily remedied if the SSIS dataset uses the same concordance relationships as SEC.

Conclusion

28. It's important to note that the US was chosen as the case study merely because of the considerable availability of detailed size-class information on the US Census Bureau website. Moreover it's equally important to state that the study was merely for the OECD SBS team to establish techniques for what could be done if countries did decide, together with the OECD, that this approach would deliver better quality SEC and SSIS data, and not to replace the current US data on the SEC database; which for 2002 onwards is already very detailed.

---

[7] Data for 2002 is however much more detailed than for earlier years.

[8] This is because a number of 4 digit NAICS groups correspond to more than one ISIC 4-digit group.

29. However the study has revealed that it is possible to improve the quality of data available on the SEC and possibly SSIS datasets by adopting this approach for countries that do not have the resources to provide information at the necessary level of ISIC detail.

Recommendation

30. In this regard the OECD SBS team plans to explore with those countries that cannot currently provide detailed data to the OECD, whether the data-transmission could be improved using this approach. As a first step in this process, countries will be asked to provide information relating to their classification system and the concordance relationships they currently use, together with inventories describing the information each country holds on SEC and SSIS related statistics (on their own classification system).

31. It's important to note that this is not the first time this issue has been raised. SBSNet discussions and a thorough OECD investigation through consultancy work on the US Census Bureau's practice have confirmed that a fundamental step to improving and facilitating data collection is the **automatisation of classification conversion routines**. In this regard OECD and individual Non-EU member countries will need to work together with their respective IT – teams to automate routines which would allow national data providers to automatically re-format and convert requested data from national data to ISIC-based data useable by OECD (see also DSTI/EAS/IND/SWP(2004)4).

32. Finally one other area of data-transmission that the OECD SBS team will be pursuing, bilaterally with Eurostat is the use of STADIUM for data transmission. For non-EU countries the SBS team will investigate whether similar approaches to data transmission could be pursued.

33. In this context data the OECD SBS team aims to engage with Non-EU OECD NSOs (or other relevant bodies) to encourage them to transmit data as part of their regular yearly activities and not as additional work. Achieving this would deliver a more symmetrical situation with EU countries that deliver data to Eurostat on a regular, transparent and agreed basis. A precondition of this exercise would be to launch a regular exchange of views and subsequent agreement on the scope and coverage of data collection activities with OECD's Non-EU countries.