# Classifications *Newsletter*

## Implementation of ISIC Rev. 4 and CPC Ver. 2 discussed in regional workshop in Panama City, Panama

A workshop on International Economic and Social Classifications was held in Panama City, Panama, on 4-8 June 2007. The workshop was co-organized by the United Nations Statistics Division and the United Nations Economic Commission for Latin America and the Caribbean (ECLAC). Substantial financial support was provided by the Caribbean Regional Technical Assistance Centre (CARTAC).

The workshop was attended by representatives from 21 countries of the ECLAC region. As organizers, two representatives from the United Nations Statistics Division were present, as well as two representatives from the United Nations Economic Commission for Latin America and the Caribbean and one representative from CARTAC. In total, thirty-two participants attended the workshop.

The workshop was the continuation of a series of workshops on economic classifications held in the region, starting in Santiago, Chile in 2003 and followed by one held in Montevideo, Uruguay in 2005. A main objective was to present, analyze and discuss the final versions of the new revisions of the international activity and product classifications, the International Standard Industrial Classification of All Economic Activities (ISIC Rev. 4), and the Central Product Classification (CPC Ver. 2), as well as the International Standard Classification of Occupations, ISCO-08. Other objectives were to review current implementation and use of economic classifications in the region, discuss classification practices in the participating countries, consider open questions in the revised classifications and plan ahead for the implementation process of these revisions in the region.

The United Nations Statistics Division gave a number of presentations concerning international work in classifications in general and the newly revised ISIC Rev. 4 and CPC Ver. 2 in particular.

Changes in concepts and rules for these two classifications were laid out and important areas of revision in the structure were highlighted. The new version of ISCO, ISCO-08, was also presented.

The United Nations Economic Commission for Latin America and the Caribbean presented information on their work programme, their web pages and data which they disseminate. Several participating countries gave presentations on the classifications portions of their work programmes as well as information on the organization of their statistical offices, which again served as a basis for discussion on practices across the region.

In addition, a number of question and answer sessions and discussions were dedicated to topics of particular interest to the participants. In this regard, the meeting addressed the question of differentiating between wholesale and retail activities. It was noted that practices vary by country and by specific case and that various criteria are employed in the decision-making, including: "type of product", "manner of marketing" and "kind of customer". UNSD made it clear that in ISIC, the distinction is made by the "type of customer (industrial or reseller, vs. household)" that the vendor primarily targets.

A special session was also set up to discuss the treatment of outsourcing, that is, activities on a fee or contract basis. ISIC provides guidelines on how to deal with specific cases, in particular where all functions of a manufacturing company have been outsourced to subcontractors. The position of ISIC in such situations is that the company is classified in manufacturing if it has ownership of the input materials. In situations where it does not have ownership of the materials, it should be classified as a wholesaler. As the practical aspects of implementing this concept can be more complex, several cases involving international borders were discussed with the workshop participants. Other issues that were addressed during the question and answer sessions included: treatment of multiple activities, "agro-industry", mixed farming, the informal sector – which remains important to a number of countries - waste

products, business coding systems, mixed occupations, support services, jobs with multiple characteristics, and timeshares.

In focussing on the implementation process for the new revisions of the classifications, it became apparent that the state of readiness and capacity to effect the necessary changes varied greatly among the countries represented. Three main points of concern were identified:

- Participants stressed the need for having classifications materials available in Spanish. This applies in particular to the classifications documents, i.e. introduction, structure and explanatory notes of ISIC Rev.4 and CPC Ver.2, which need to be provided as soon as possible in Spanish to allow for the timely implementation of these classifications in the region. Participants suggested that the existing translations of draft versions, carried out by some countries, should be used to speed up the process of official translation.

- Participants noted the need for a dictionary of classification items that would contain detailed descriptions. This was recognized as an important tool in ensuring that technical classifications terminology would be correctly understood.

- Participants suggested that a structured set of questions be prepared that can be used to facilitate the classification of activities in ISIC Rev.4. This will be important for the application of ISIC in business register coding as well as in household surveys, where the interviewer cannot always be fully aware of all ISIC details and coding options.

# When is a category in the international reference classification significant for a national activity classification – some practical suggestions

*Matthias Greulich*
*Federal Statistics Office of Germany*

In the global setting, statistical data for different countries or economic areas must be made easily comparable if they are to serve as a basis for political and business decisions. Among other things, this requires national activity classifications to be directly comparable to or consistent with the international reference classification(s). This does not necessarily mean that the national activity classification must follow the structure of the reference classification

strictly. It is enough if their elements can be clearly allocated, in a mutually exclusive way, to the elements of the reference classification. Of course it is also necessary that the methodological concepts of the reference classification be respected at national level.

Based on the given structure of the reference classification, every country may wish to define elements (categories) for a national activity classification, which meet their national needs. This may require a compression of categories of the reference classification, if a certain activity does not exist in a country or economic area or if the activity is insignificant. An example can be the three-digit category for sea and coastal water transport in ISIC Rev. 4, which would not need to be disaggregated for a landlocked country. On the other hand it might be necessary to expand the breakdown of the reference classification at national level. The question is: how can decisions regarding a national breakdown be made?

A conference held in Williamsburg, U.S.A in 1991, dealt with this issue.[1] A formal approach for determining candidate categories to expand or compress in the reference classification can be an analysis of their relative importance as measured by: the number of statistical units engaged in a certain activity, the value added or turnover resulting from the activity, the number of employees carrying out the activity, etc. In this case the relation between the value of a certain variable for the category in question and the average value for all categories within the classification item on the next (higher) hierarchical level would be calculated for each variable. As an example this could be the number of statistical units of an ISIC class in relation to the average number of statistical units of all classes of the group. If the relation is (for example) between 0.5 and 1.5, the category could be kept or created respectively. If the relation is lower, the category of the reference classification could be merged with another category. If the relation is higher, a further split could be appropriate. An advanced approach could be the calculation of weighted ratios for a combination of economic variables.

1. U.S. Department of Commerce, Bureau of Economic Analysis, Economic Classification Policy Committee. Issues Paper No. 4, "Criteria for Determining Industries" Washington D.C., U.S.A, October 1993.

In the development of a classification to be implemented in the future, such an analysis must necessarily be based on estimated figures and/or figures from non-statistical sources. In every case the decisive thresholds (in the above-mentioned theoretical example, 0.5 or 1.5 respectively) must be predefined and (for the weighted ratios) the weights for the different variables must be set. These are undeniable disadvantages in the approach described above. Furthermore, it is a fact that the level of detail already in existence has an impact on the probability of exceeding a given threshold. In an area of the classification which is already very detailed, it is easier to exceed a certain threshold (and therefore to keep or to create a category), than in a less detailed area of the classification. What is more, this formal approach ignores dynamic aspects, such as the question of whether an industry is growing or diminishing.

Another possible approach for determining candidate categories to compress or expand is the calculation of homogeneity ratios as described in the Introduction to ISIC Rev. 3. A disadvantage is that there is obviously no definitive and mutually exclusive description of activities by products in the world and not enough usable data for their calculation. This became clear again during the meeting of the Technical Subgroup to the Expert Group on International Economic and Social Classifications which was held in April 2007, where the concept of homogeneity ratios was critically discussed.

In Germany a more pragmatic approach was used for developing a national activity classification based on ISIC Rev. 4 or the European activity classification, NACE Rev. 2, respectively. Central ideas in this approach were:
 • A national subclass should only be created if a need is expressed by data users and if the intent is to use it for data collection;
 • When creating a national subclass, possible confidentiality problems (for the proposed new subclass but also for the remaining residual subclass) must be taken into account in advance;
 • The extra burden for respondents and statistical offices resulting from the addition of subclasses should be taken into account.

Data users within the statistical system (e.g. National Accounts) and from external institutions, such as ministries, professional associations and research institutes were invited to express their needs for a further split of the reference classification. A clear justification was needed for any proposal. Additionally, estimations of the number of statistical units with their principal activity in the proposed new subclass and of the expected turnover for such a class were compulsory requirements. No strict threshold was defined in terms of statistical units or turnover for new subclasses, in order to take growing industries (such as solar technology), or well justified user needs, into account. However, attention was paid to the necessity of avoiding data confidentiality problems at national and regional levels. As a result, not all proposals could be implemented.

When necessary, the justification of a given proposal and the above-mentioned estimations were questioned in a written consultation and in bilateral meetings. The final draft was noted with approval by a consultancy committee composed of representatives of data users and statisticians. This exercise was indeed a big challenge and did not proceed without some intense discussion. However, it finally lead to a generally accepted national version of ISIC Rev. 4 or NACE Rev. 2 respectively, which will be the basis for statistics in the coming years.

# General problems in applying ISIC in population censuses[2]
*Eivind Hoffmann*
*Head, Statistics and Analysis Division*
*Norwegian Directorate of Immigration*

This note outlines the general problems in applying ISIC or any other activity classification in household based data collections, such as population censuses. It presents the different strategies and preparations that are necessary to ensure effective and reliable coding of the industry in such data collections, as well as the objectives and main strategic choices to be

2 This note is based on Chapter Four "Coding of Occupations and Industry", *Collection of Economic Characteristics in Population Censuses*, (ST/ESA/STAT/119), United Nations and International Labour Office (2002), New York and Geneva. Peter Elias, Brian Embury, Eivind Hoffmann, David Hunter and Roger Thomas contributed to that text.

made, with the main organizational factors determining the effectiveness and success of the coding and processing tasks.

Why include an activity classification in a population census, given that obtaining reliable results will be complicated and costly?   The short answer is that a population census in most countries will be the only possible source for statistics that will reflect employment in all economic units in the whole country,[3] to a significant degree of detail on both industry structure and on their geographic distribution.

When applying ISIC in a population census the aim is to determine and correctly record the groups to which the respondents' jobs belong in the industry classification, at the most detailed level of the classification possible, on the basis of the information provided in the response.   This task has to be completed within an overall processing plan for the census, to a pre-specified timetable and either within pre-specified cost limits or in a fashion which will minimize cost, given the level of statistical detail required to satisfy important user needs in the local and national context, as well as for international reporting.

The precise impacts of these aspects will depend on the choices made with respect to some strategic aspects of the coding and processing of census forms, and which are discussed in the following section.

### Strategic coding and processing options
In population censuses there is a choice among:
(a)   the respondent or the enumerator selecting a predefined group;
(b)   the enumerator coding in the field, either during the interview or before the responses are forwarded for further processing;
(c)   coding done by specially trained coders in connection with data entry.

It is important to note that only strategy (c) will provide reliable statistics for industry groups defined

---

3   Compared with the scope of the national accounts, the main exclusions in a population census are the persons commuting across the external borders to work in the country, and the main inclusions are persons commuting to work outside the country. Persons living in institutional households will normally be excluded from the scope of labour force surveys, but not from a population census.

at a level of detail corresponding to two or three digits in ISIC Rev. 4 and that experience clearly shows that the industry responses recorded in a population census are very uneven in the level of detail they will provide. Many responses will support detailed coding, especially if the questions are formulated along the lines outlined below.  At the same time a significant number of responses will not support a predefined level in the structure of ISIC. By insisting on a predefined level, the coding process may therefore both lead to unnecessary loss of information for a large part of the returns and to misrepresentation of the data quality for other parts.

### Coding tools
Given well formulated questions (see below), designed to obtain information on the type of activity at the work place, the key to success in applying ISIC in a population census is that good coding tools had to have been developed and that coders and their supervisors had to have been well trained in their use. The basic tools required by coders will include:
(a)   Coding instructions: These should cover all operations that the coder is required to perform;
(b)   Coding index: This is the key coding document through which verbatim terms incorporated in job titles, descriptions of tasks, etc., are translated into codes;
(c)   Query resolution procedures: A query occurs when coding clerks cannot assign a code using the specified procedures and indexes. There should be clear instructions on when and how the coders should raise queries, how to record and report them, and their resolution. Queries are the most useful inputs to both immediate and future work in updating the coding index and the classification itself.

### Use of automatic or computer assisted coding
Systems for automatic or computer-assisted coding (CAC) of industry responses are used in a number of countries. The introduction of these systems may well have had beneficial effects on the consistency of coding. They may significantly reduce the combined coder and data entry working time and the time required to complete the task if it is possible to combine the systems with automatic (optical) reading of the free text written responses or digital imaging of all the completed forms. Such techniques may significantly reduce the task of transcribing the verbatim material into a computer-readable form, facilitate the reading of the responses on screen, and make possible the entry of only a truncated version of

the responses for the CAC system to match against a (good) coding index.

A CAC system that obtains automatic coding of 60-70 percent of all responses will represent significant gains, in particular if this is integrated into a data entry and processing system that starts with optical reading or scanning of the questionnaires.

**Formulating questions on establishments and activities and recording replies**

The industry question(s) should have two components. The first component should ask for the name and location of the place of work and request details of division or branch for larger private enterprises and government institutions and establishments. Asking for the full address may also be useful, in particular, when access can be had to an establishment register. Unfortunately few countries have a comprehensive updated register of employers with separate identification of all establishment locations. Thus, in most countries the detailed name and address of the establishment can only be used as a source for correct industry code for some of the respondents, and a description of the activity at the establishment has to be used as the basis for assigning codes for the rest. The second component of the industry question(s) therefore should be to request a good description of the main products or the main functions of the place of work.

The industry code should refer to the establishment where the person actually works, and not the legal unit to which the establishment may belong. For example, a major company in a country, 'XYZ Ltd', may be well known but engage in a variety of activities at different locations around the country. Generally industry should be coded for each separate establishment (usually at separate locations), not only to the main activity of XYZ Ltd. With access to a reasonably up-to-date register of establishments, where industry is coded for each unit, the coding can be done by reference to the register and be compatible with employer surveys based on the same register. For example, XYZ Ltd may be primarily an iron and steel mill but may also have XYZ Coal Mines, XYZ Iron Ore Mines, XYZ Health Clinic and XYZ Bus Services as separate establishments. Entries in the coding index for industry reflecting this structure will help to ensure that respondents working at these separate establishments are given their relevant industry codes and not simply coded to the code

given to XYZ Ltd. A similar listing of major government departments and organizations is also very useful in coding respondents working for the central or local government to their correct industry codes: for example coding respondents working with the municipal health administration separately from those working in municipal hospitals.

If census and survey planners feel forced to use only one question for industry, due to space and time limitations for the individual interview, it will still be better to ask a combined question on the name, location and type of product, rather than only asking about the production at the place of work.

**What is a coding index?**

The process of coding industry information involves the task of matching responses to index entries to find the appropriate codes. The coding index is the key instrument for this matching process, linking the words used in the various parts of the response to the numerical code that represents the allocation of that response to the corresponding group of the classification. The coding index guides the coder by listing information, e.g., key words, which can be found in the responses, and indicates how different responses are allocated to the detailed or more aggregate groups of the classification, depending on the nature of the information in the response and on the instructions for the coding process. The collection and coding of the elements to be included in the coding indexes has to be carried out by experts on the industry classification to ensure that they are done correctly. The work will be painstaking and time-consuming, but the investment involved in the collection and coding of up to 5, 10, 20 or even 30 thousand index entries will prove well worth the effort in terms of the speed and reliability with which hundreds of thousands or millions of census forms can be coded during a census operation.

The advantages of having a well-structured coding index are threefold. Firstly, it causes the coder to search for index entries in a way that is consistent with the coding rules. Secondly, it speeds up the task of coding by restricting the coder's search through the index because of the smaller number of entries. Thirdly, when the index is searched, either by computer or a coder, it reduces the risk of finding matches with words that are irrelevant for the coding decision, thereby reducing error.

If using two different forms of questions as the basis for determining the industry and the respondent's place of work, most census and survey coding operations will find it useful to have two "coding indexes" for the coding of industry:

(a) a list of as many as possible of the establishments which are/were operational in the geographic region covered by the coding operation, where each establishment has been given the correct industry code by those who are specialists in establishment surveys and the coding of establishments' activities. In practice such lists (or business registers) may often cover only large, formal sector establishments as they have been created from lists kept in tax offices, licensing offices and/or chambers of industry and commerce. They may nevertheless cover significant portions of the work force, and their use for coding will eliminate one possible source of inconsistency in employment statistics between census results and the results of establishment surveys.

(b) a list of significant word combinations reflecting the answers given and recorded in response to the questions about the type of product or service that is being produced at the place of work or the functions of the given place.

The industry coding process will therefore usually involve firstly an attempt to match the name and address of the person's employer with those in the list or register of establishments. If a match cannot be made using the register of establishments, then an attempt is made to match the description of the industry with the index of type (b) above.

## Calendar of Meetings

**Technical Subgroup Meeting**

The next Meeting of the Technical Subgroup of the Expert Group on International Economic and Social Classifications is tentatively scheduled for 4-8 February 2008. This meeting will focus on finalization of the Companion Guide to ISIC and CPC which is being developed as one of the tools for implementation of ISIC Rev. 4 and CPC Ver. 2.

## Editorial note

The Classifications Newsletter summarizes recent developments in the field of international classifications, announces upcoming events and draws attention to the availability of relevant classifications material in print and on the Internet. The Classifications Newsletter can be found at the United Nations Classifications Website http://unstats.un.org/unsd/class under "Newsletter".

To receive the Classifications Newsletter by e-mail, you can sign up for our mailing list at http://unstats.un.org/unsd/cr/registry/regmaillist.asp or select "Mailing list" from the menu at the Classifications Website. For further information please contact the Classifications Hotline:

**United Nations Statistics Division**
**Attn: Economic Statistics and Classifications Section**
**United Nations**
**New York, NY 10017, USA**

**E-mail address: chl@un.org**
**Fax: 1 212 963 1374**

**Classifications Website:**
http://unstats.un.org/unsd/class