

Methodological issues
in the development, use, maintenance and revision
of statistical classifications

Introduction

1. The purpose of this note is to outline some of the main methodological issues which have to be addressed in the development, use, maintenance and revision of the classifications which are being used when producing statistics. The discussion is intended to be general, but the examples used as illustrations will mostly be taken from work with occupational classifications in general and the *International Standard Classification of Occupations (ISCO-88)* in particular, as these are what I know best.
2. By "a statistical classification" I shall understand a set of discrete values which can be assigned to specific variables which are to be measured in a statistical survey, or registered in administrative files which will be used as basis for statistics.
3. Such value-sets are frequently not constructed by the statisticians but by persons responsible for administering public policies (e.g. custom regulations, criminal legislation), providing public services (e.g. job placement, education, health services) or analysing social, economic or natural phenomena. When adopting and adapting such value-sets for their use in statistical inquiries the statisticians must, however, understand whether and how the various methodological issues have been tackled as well as contribute to finding improved solutions to them wherever necessary and possible.

Developing the classification

4. We can distinguish three types of methodological issues which have to be addressed in the development of classifications: (a) users' requirements issues, (b) conceptual issues and (c) collection of information issues.
5. The main users' requirements issues are:
 - (a) how to determine who the main users are;
 - (b) how to determine what the different users would like the classification to do;
 - (c) how to balance the different users' requirements against each other or make a choice between them when they are contradictory?

6. Given the users' requirements the main conceptual issues are:

- (d) What are the (main) variable(s) for which the value-set is valid? (The "variable" to which ISCO-88 is to be applied is "main tasks and duties" of work performed, and not other aspects of work, such as "exposure to hazardous substances" or "uncomfortable working conditions". Thus the variable "occupation" can be said to be defined as "the main tasks and duties of work performed".)
- (e) What are the primary units for which we can measure the main variable(s)? ("Jobs" are the primary units for ISCO-88.)
- (f) What are the rules for linking other units to the primary units so that these other units can be assigned values of the classification variable? ("Persons" can only be assigned an "occupational group" through their link to a past, present or future job. They can only be given an "industry" code by being employed by an establishment.)
- (g) What are the conceptual rules for identifying same value of the variable? (For ISCO-88 the rule is that "a set of jobs whose main tasks and duties are characterised by a high degree of similarity constitutes an occupation", which is the most detailed element in the value-set.)
- (h) What are the similarity criteria used to define higher level categories (aggregated value-sets) in hierarchical classifications? (In ISCO-88 the main similarity criteria are the "skill level" and "skill specialisations" needed to carry out the tasks and duties of the jobs.)

7. The main collection of information issue is

- (i) how to collect the information which describes the defined categories and the dividing lines between them. (For an occupational classification one needs to collect information about the main tasks of jobs in the different occupations for the whole range of work situations as they can be found in establishments of different size and in different industries. If one wants to be able to apply different similarity criteria to define alternative aggregated value sets, then the corresponding information must be collected.)

Using the classification

8. In this context "using the classification" means the recording of the appropriate value of the value set for a particular primary unit. For statistical surveys and administrative records the following issues are important:

- (j) Respondent, interviewer or expert coding? (With respondent coding the respondent is presented with a value-set and asked to indicate the value which best applies to him/her. With interviewer coding it is the interviewer who determines the correct value on the basis of the information received from the respondent. With expert coding the respondent or the interviewer writes down the information from which someone (the expert) later will determine the correct value.)
- (k) What question(s) to ask the respondent will depend on the choice of coder, cf. i. (The

coding respondents need to be told on what basis they should choose one of the possible categories. The interviewers need to get the information relevant for coding either by them or the experts. With interviewer or expert coding of occupation it seems that to ask questions to determine an occupational title and main tasks and duties is the best way to determine the occupational code.)

- (l) instructions about what response elements to use and how to use them are needed to ensure that interviewers record the most important response elements and that the coders use them consistently. (The coding process should correctly retain as much as possible of the information provided by the respondents. The ILO recommends that when coding "occupation" one should code to the most detailed level in the classification supported by the information provided, and that title, task and other information should be used according to clearly specified rules.)
- (m) How to develop appropriate coding tools? These issues are concerned with the development of coding indexes reflecting the type of responses which are given, as well as with the coding rules; and with the development of automated or computer assisted coding procedures which reflect the rules to be used for accurate and effective coding.
- (n) Quality control procedures are needed to monitor the quality of the coding the coding process and to provide feedback both to the coders and to those responsible for the classification and the coding tools.

Maintaining the classification

9. By maintenance of the classification we understand those activities which are needed to ensure (i) Correction of errors made in the construction of the classification (i.e. value set) and associated coding tools; and (ii) Up-dating of the descriptions of the value-sets, dividing lines and coding tools of the classification as previously unknown or genuinely new types of primary units (e.g. jobs, occupational titles) emerges or are discovered, or as new information is obtained about existing types. Maintenance should be an ongoing activity of those responsible for the classification, to be combined with the training and back-stopping of the users of the classification and of the users of the resulting statistics. The main methodological issues are:

- (o) How to best organise maintenance activities.
- (p) Criteria for identifying "significant" effects of maintenance activities on comparability of data;
- (q) the development of methods for correcting and/or smoothing statistics for effects of maintenance.

Revising the classification:

10. Revising the classification involves a complete review of users' needs as well as the conceptual basis and the users' tools, i.e. all the "solutions" arrived at for the issues outlined under a.-m. above. Such reviews should only be undertaken at long intervals (10-15 years) or if there

is compelling evidence that revisions are necessary. Such evidence may come from national users or from international developments. The activities involved in revising the classification are essentially the same as those needed to develop it. The additional methodological issues involved are:

- (r) How to determine whether a revision is needed or not?
- (s) How to determine whether new solutions are better than current practices?
- (t) How to implement a revised classification in on-going statistical programmes, given the need for comparability with past statistics?
given the need for comparability with past statistics?

Concluding remarks

11. A list of 19 different types of methodological issues for statistical classifications may seem discouragingly long. With some classifications the "resolution" of some of these issues will be trivial or self-evident. However, they need to be examined if we want to obtain statistics of high (or, at least, known) quality with the resources available. It is also my belief that good national practices with respect to these issues will contribute significantly to ensure international comparability of the statistics using the respective classifications, which also, of course, depends on how countries "map" their national classifications to the international one - an issue related to issue s. above.