

POLICY ON THE DOCUMENTATION OF DATA QUALITY AND METHODOLOGY

OVERVIEW

The National Statistics Office, as the main national institution producing official statistics, has the responsibility to inform users of the concepts and methodologies used in collecting and processing data, of the quality of the data produced and of other features that may affect the use or interpretation of statistics produced.

It is NSO policy to be as open as is reasonably possible about its statistical methodologies and to provide adequate documentation of methods and quality measures to users. This is done in order to allow users to:

- ◆ Verify and determine whether the statistics are adequate for their intended use;
- ◆ Recognise and assess the level of accuracy of the statistics under consideration.

DATA QUALITY DESCRIPTIONS

In view of the above, the NSO shall, subject to cost and flexibility constraints endeavour to provide data quality descriptions that would cover the following aspects:

- ◆ **Coverage** the quality of the survey frame or list (for surveys or censuses) or source files (for administrative data) as a proxy for the desired universe should be addressed (including gaps, duplications and definitional problems).
- ◆ **Sampling error** if a survey is based on a random sample then estimates of the sampling error of tabulated data based on the sample should be provided, together with an explanation of how these standard error figures should be used to interpret the data. The method of presentation may vary from explicit estimates of sampling error to use of generalised tables, graphs or other indicators. If the survey is based on a non-random sample, the implications of this on inferences that might be made from the survey should be stated.
- ◆ **Response rates** the percentage of the target sample or population from which responses or usable data were obtained (on a question-by-question basis if appropriate) should be provided. Any known differences in the characteristics of respondents and non-respondents should also be described, as well as a brief indication of the method of imputation or estimation used to compensate for non-response.
- ◆ **Comparability over time** it may be appropriate to discuss comparability to the results of the same activity for a previous reference period, especially if there has been a change in methodology, concepts or definitions. If such a change is thought to affect comparability between one time period and another, a quantitative estimate of this effect should be made whenever possible.
- ◆ **Benchmarking and revisions** the experience acquired by the NSO should be translated into guidance on the possible impact of benchmarking or revisions or comparability over time.
- ◆ **Comparability to other data sources** if similar data from other sources exist they should be identified. Where appropriate, a reconciliation should be attempted describing how the various data sets differ and the reasons for these differences. Comments on the quality of the various data sets should be provided if an evaluation is available.
- ◆ **Other important aspects** there may be other aspects of data quality that are of prime importance given the objectives of a specific activity. These should be included with the basic indicators of data quality. Examples are: unusual collection problems; misunderstandings of intended concepts by respondents; incidents such as disasters and strikes which may impact on key variables; changes in classification or in its application; and response based on financial years that do not correspond to a fixed reference period. In different surveys and at different levels of aggregation, different sources of error may predominate. Subject to cost limitations, the most important sources of error should be evaluated periodically, and the results made available to users in an appropriate form.
- ◆ **Non-response bias** an assessment of the effect of non-response on the collected statistics should be provided where possible.

- ♦ **Response bias** evidence of response bias problems resulting from respondent misunderstanding, questionnaire problems, or other bias sources, should be provided when available.
- ♦ **Editing and imputation effect** the effect of editing and imputation on the quality of data should be assessed.
- ♦ **Any other error sources** if there are particular sources of error or unforeseen events that are relevant to the series or occasion, these should be noted and described.

Guidelines on the Description of Methodology

For data resulting from censuses or surveys, the methodology reports should provide at least an outline of the main steps in conducting the census or survey. They should also provide more detailed information on those aspects of census or survey methodology which have a direct impact on the quality and applicability of the data produced from the respective census or survey. The following aspects should be covered where applicable:

1. Objective of the census or survey;
2. In the case of surveys, the target universe and any differences between this and the survey frame actually used;
3. The questionnaire(s) used and all important concepts and definitions;
4. The method used for data collection (e.g. telephone, mail or personal interview and details of any follow-up procedures for non-respondents);
5. Any manual processing (e.g. coding) that takes place prior to data capture;
6. The method of data capture;
7. Quality control procedures used in connection with operations 4-6 above;
8. Procedures for editing the data and for handling non-response and invalid data;
9. Benchmarking and revision procedures where used;
10. Seasonal adjustment methods used;
11. The form in which the final data is stored and the tabulation or retrieval/query system, including confidentiality protection procedures;
12. A brief summary of the results of any evaluation programmes;
13. Any special procedures or steps that might be relevant to users of the survey data.

Methodological Descriptions of Data emanating from Administrative Sources

For administrative data, or data partially emanating from sources outside the NSO, it is good practice to provide the background information outlined in the previous section. As a minimum, the following topics should be covered:

- ♦ Data sources;
- ♦ Purposes for which the data were originally collected;
- ♦ Merits and shortcomings of the data for the statistical purposes for which they are being used (e.g. in terms of conceptual and coverage bias);
- ♦ Description of the data process following receipt of data; any corrective measures to the original data sets;

Reliability of the estimates, including any caveats, as appropriate.