ACC SUBCOMMITTEE ON STATISTICAL ACTIVITIES SA/2000/10/Add.4
Thirty-fourth session 6 September 2000
Washington, D.C., 20-22 September 2000
Item 12 of the provisional agenda

QUALITY ASSURANCE, ARRANGEMENTS WITH COMMON QUESTIONNAIRES AND CONFLICTING STATISTICS

Note by Food and Agriculture Organization of the United Nations[1], Rome

The FAO databases are being maintained not only to facilitate dissemination of global statistics pertaining to food and agriculture (including forestry and fishery) through yearbooks (and on the internet) but also to enable the review and monitoring of the progress of nations through documents such as the State of Food and Agriculture (SOFA), the State of World Fisheries and Aquaculture (SOFIA), the State of World's Forest (SOFO), the State of Food Insecurity in the World (SOFI), etc. As a consequence, the databases are in the form of an annual time series[2] covering practically all countries of the world. Furthermore, these time series do not refer to only the primary or basic statistics but also a number of derived statistics such as index numbers of food and agricultural production, estimates of food availability (obtained through the preparation of food balance sheets), etc.

As regards the compilation of the primary data, FAO, as in the case of other international organizations, relies on a set of questionnaires that are regularly sent to countries. However, there are a number of countries where the available (official) data are incomplete in terms of (a) the range of variables covered (for example in many countries data on items of agricultural output are limited to important items and data on agricultural inputs are practically non-existent), and (b) coverage of the nation (sometimes parts of the country are excluded from the statistical reporting system). [3] Most of these countries are characterised by poor resources and weak infrastructure. Even after the years of efforts, the importance of information for decision making has not been given higher priority in these countries. Thus in order to complete and up-date the different time series, the data gaps, resulting from incomplete or non-reporting by countries or reports which are considered unreliable, need to be filled by FAO estimates. As a consequence, the primary data in the FAO databases do not include only the information returned by countries through questionnaires but also the estimates introduced by FAO. Keeping in view these features of the FAO databases, the issues relating to quality assurance etc., are presented below.

(a) Quality assurance

---

[1] Prepared in the Statistics Division, Economic and Social Department. Comments and suggestions from Forestry and Fisheries Departments are gratefully acknowledged.

[2] Containing relevant data for each variable in the domain. However, point estimates for certain variables like forest cover, are also prepared.

[3] For details, please see Issues and Concerns for Developing Countries by Kabat, L. et al. - Paper submitted to International Conference on Agricultural Statistics, Washington D. C. (1998).

FAO generally places emphasis on the "official" nature of the national sources reporting the data through the annual questionnaires. Nevertheless, the reported data are thoroughly reviewed and analysed and adjusted where necessary before publication.[4] Where there are gaps, national publications along with other reports are used to complete the series.[5] In this context the process of assurance of quality involve the following considerations:

1. *Relevance of statistical concepts*: The basic concepts have originated from traditional agricultural censuses and surveys conducted in various member nations over the years. However, emphasis is being placed on harmonising these concepts with the concepts adopted (or recommended) by other international systems.

2. *Comparability of statistics*: Quality of data depends partly on how well the basic ratios or averages can be estimated for making international comparisons and presenting the world and regional picture. To meet this basic need it is necessary to examine the comparability of the data over space and time. As there may be differences between the concepts or definitions used by a nation and those given by the FAO, care is taken in adopting the national data for making valid regional analysis and international comparisons. Where necessary the national data are either adjusted for differences in the concepts over space and time or alternatively the differences are explained and quantified by giving adequate notes.

3. *Accuracy*: The degree of accuracy of the data is different in different countries. In the case of the FAO estimates, there is neither any statistical measure (like standard error) nor any reference population total to determine the accuracy of the data sets. However, internal analyses are carried out taking into account: (a) the latest year for which official data are available, (b) the extent of revision made in subsequent years, and (c) share of data based on official estimates, in making a final judgement about the accuracy of a series. [6] Country missions are also undertaken to improve the quality of the data.

(b) Arrangements with common questionnaires and the policies for publishing the data collected in this way by the organisations involved

With regard to agricultural statistics, the other organizations involved are OECD and Eurostat which compile data for the developed or industrialized countries. In view of the basic features of the related databases (especially the degree of details for each item of the data-sets for which information is required) a great deal of efforts are required for harmonising the concepts and definitions used for collection of

[4] In so far as production and trade of agricultural commodities are concerned the review and analysis are undertaken within the framework of supply/utilization accounts.

[5] A variety of methods are being used by the FAO for preparing estimates for missing observations in a time series. If no suitable variable has been found to make a regression/ratio estimate a method that is frequently used is to interpolate or extrapolate using a linear trend. Estimates are also made on the basis of available information, for example the arable land is often estimated by adding cultivated temporary crops in pure stands (without double or multiple cropping) + the fallow land (if known or estimated on the basis of last official figure) or by adding temporary crops harvested area (gross cropped area) and deriving the net crop area with the help of cropping intensity. Cropping intensity is estimated on the basis of crop area for which both cultivated and harvested area is known for the years for which data are available. However, sometimes it becomes unavoidable to repeat the data for the last year.

[6] Another kind of analysis is the calculation of outliers and inliers to investigate into anomalies and identify possible errors. A commonly used approach is to look for special events to check the accuracy of the data supplied by national organisations. This effort is supplemented by evidences found in the literature about future policies, for examples a decrease of crop production and the intention to divert money to other fields, agreements between country to decrease cultivated land, etc. wherever possible.

primary data before the issue of single common questionnaires can be considered. [7] Accordingly the progress achieved is different for different data-sets. In the case of crop, livestock and fisheries at present, there are some ECE/FAO/OECD/Eurostat and other inter-agency fora[8] where these issues are under active consideration. Even if joint questionnaires are agreed upon, it is understandable that each organization will publish/disseminate the data in the most appropriate manner to suit its requirements.

However, in the context of forestry statistics, considerable progress has been made towards common questionnaires. The Forestry Department is currently executing for the second time a Joint Forest Sector Questionnaire with Eurostat, ECE and the International Tropical Timber Organization (ITTO). This builds upon more than ten years of a joint forest products questionnaires executed by FAO, ECE and Eurostat. However, each of the organizations publishes the data independently according to their specific needs.

(c) Conflicting data in organisations even when the same original national data is used as a starting point

It is not surprising to get conflicting data on an identical subject when these are released by different Organisations at different places. Some of the common causes for this situation are given below:

1. Different practices relating to reference period, time when the data are supplied by the country, source (reporting office) etc.
2. Specification provided (or required) by the organisation collecting the data.[9].
3. Conceptual differences in the data collection procedure[10]
4. Method of estimation[11] used by individual Organisations for completing missing observations.

(d) Conflicts between nationally supplied data and internationally comparable data prepared by some international organisations.

Conflicts between nationally supplied data and those published by international organisations are due to differences in concepts and classifications used by various countries arising out of the fact that the national statistical systems are based on local administrative needs and conventional practices. The

[7] For example see "Handbook of Concepts and definitions used in international collections of food and agriculture statistics" issued jointly by UNECE, Eurostat, FAO and OECD in 1995.
[8] Like Intersecretariat Working Group on Agricultural Statistics (IWG.AGRI) and the Coordinating Working Party on Fishery Statistics (CWP).
[9] This situation is best illustrated by the producer prices data. For example, the EUROSTAT requires information on the producer prices by variety of different products in their questionnaire while the FAO collects data on the national average producer prices for individual commodities.
[10] For example, details of irrigated area by source (given by a country) might have been derived using either as the area equipped to provide water to crops (by adding areas under full and partial control irrigation, spate irrigation, equipped wetland) or as the total water managed area (i.e. the area equipped to provide water to crops + other irrigated land areas which are not equipped with water control structures - wetland, flood recession cropping area)
[11] See footnote 4.

situation can best be studied by comparing[12] many statistics like the land use statistics, pesticides consumption, being published by FAO and the data supplied by some countries. [13]

[12] For example, let us consider the USA data on pesticide consumption. The USA classification of Pesticide Usage in Agriculture differs from FAO's. For instance, plant growth regulators (which are classified as a pesticide by the FAO) are included under "herbicides". Similarly USA includes "mineral oils" in the other pesticides and "Molluscicides" (which is classified under "Insecticides" by the FAO) is included under Rodenticides.

[13] A variety of methods are being used by the FAO for preparing estimates for missing observations in a time series. One of the most common methods is to compile 'Supply and Use Account' for individual primary agricultural product. Another method frequently used is to compile linear trend against the time if no suitable variable has been found to make a regression/ratio estimate. Estimates are also made on the basis of available information, for example the arable land is often estimated by adding cultivated temporary crops in pure stands (without double or multiple cropping) + the fallow land (if known or estimated on the basis of last official figure) or by adding temporary crops harvested area (gross cropped area) and deriving the net crop area with the help of cropping intensity. Cropping intensity is estimated on the basis of crop area for which both cultivated and harvested area is known for the years for which data are available. However, sometimes it becomes unavoidable to repeat the data for the last year.