47918

Distr.: LIMITED

ECA/DISD/STAT/IQAS/98/6
9 December 1998

Original: ENGLISH

UNITED NATIONS

ECONOMIC AND SOCIAL COUNCIL

ECONOMIC COMMISSION FOR AFRICA

Regional Workshop on improving the quality of
African statistics for English-speaking countries

Addis Ababa
14 - 18 December 1998

Data Integration

TABLE OF CONTENTS

DATA INTEGRATION

I. Introduction

(a) Objectives

1. The prime objective of data integration is to achieve a coherent and well coordinated system of official statistics. A second objective is to optimize the use of limited resources while at the same time minimizing statistical discrepancies and improving the reliability of data through more consistency checks which are made possible by integration.

(b) Concept of integration

2. Integration, as a concept, arises from the fact that organizations have what are called the vertical dimension and the horizontal dimension. The vertical dimension refers to the hierarchy of the various organizational units (divisions, branches, etc.) and how they relate to each other with respect to their functions and activities. The horizontal dimension of the statistical agency refers, for each level, to the across-the-board interdivisional and inter-branch activities. It deals mainly with the integration of outputs, and is invisible from outside since it does not show on the organizational chart. It is a function which is very critical to a statistical service in view of the complex nature of its outputs.

(c) General considerations

3. Integration does not simply mean implementation of activities in the various divisions, sections and units. It also includes the overall philosophy of the organization. Barriers between various parts of the organization must be overcome, and a multidisciplinary approach to implementation of activities should be fostered. For statistical activities, integration ensures that concepts, methods, classifications and definitions do not exist in isolation but are fully unified with the overall philosophy and operations of the statistical service.

4. Integration poses a great challenge to the management of the statistical organization. In the centralized statistical system, the responsibility for integration rests with the Central Statistical Office and for this internal committees could ensure a certain degree of integration. In a decentralized statistical system, the need for both internal and external committees becomes even more critical for coordinated introduction and use of the various integration instruments. This issue of inter-office and inter-agency coordination is discussed further in the section on institutional arrangements for data integration.

5. Integration in household surveys means the "use of same concepts, survey personnel, facilities, sampling frames and related materials in multiple surveys and survey rounds" (UN, 1986). Through this approach it is expected to gain efficiency and quality of data produced which cannot be achieced in stand alone operations which are not linked. The linkage of the surveys has the advantage of reducing overall costs of the survey programme and value of the survey results is enhanced. Surveys

which are based on this approach are usually classified "Integrated Household Surveys".

II. Data integrators and integrating frameworks

6. As already noted, coherence is a necessary aspect of the quality of statistical information. In a system of official statistics, the data collected must be meaningful, comparable and capable of being used in or be related to other fields. In this context, the ECA has in the past identified three main fields of official statistics: (a) economic statistics, (b) social statistics and (c) environment statistics. The Strategy for the Implementation of the Addis Ababa Plan of Action for the Statistical Development of Africa in the 1990s schematically presents these relationships (see Annex: Fields of statistics and their interrelationships). Inter-relationships in data collected is one of the most critical elements for the implementation of an integration programme.

7. Methods used for ensuring that statistical data are coherent, comparable and relate across fields include, as discussed above, use of multi-purpose household surveys, suitable sampling frames and permanent field survey facilities. The population and housing census, as already noted is also a strategic tool for ensuring coverage of data collection programmes, as well as in providing bench marks for a wide variety of data. In addition, the following tools: common classifications, bridge-tables for linking classifications, clear and precise identification of statistical units, common concepts and definitions, establishment registers, and integrating frameworks such as those made possible by the System of National Accounts.

(a) Population and housing censuses

8. The Population and Housing Censuses are a basic means for establishing an integrated statistical data collection programme. The population and housing census is "aimed at providing a comprehensive source of statistical information for economic and social development planning, for administrative purposes, for assessing conditions in human settlements, for research and for commercial and other purposes" (UN, 1998). The population and housing census is an important integrating instrument on account its essential features, namely individual enumeration, universality within an individual territory, simultaneity and defined periodicity. In addition to the enumeration of the population, the population census can also be used to collect economic information about the population, for example, data on industry and status of the economically active population. Such data would indicate if the person enumerated is an employer, an employee or an own account worker. It would also indicate the address and physical location of unincorporated household enterprises. These data are used to compile the different accounts of the household sector of the national accounts. Further, the information can also be used, in establishing a register of small enterprises. The register can be integrated with other data to establish a universe of all enterprises or establishments. Population census data and economic survey data may thus be integrated to measure various aspects of an economy's structure and performance and so provide useful data for policy and decision making purposes.

(b) Statistical units, classifications and concepts

9. Statistical units are important elements of a data integration programme they are the primary units of observation and for recording of data sometimes referred to as micro-data. Apart from population censuses, the statistical units for most surveys are the (a) establishment and (b) the household. The establishment is the unit for which a range of homogenous and single location production related data are collected. For agricultural establishments the single holding is taken as equivalent to the establishment. This is the same concept as was used for the 1990 World Census of Agriculture and is also used in the 1993 System of National Accounts for the analysis of economic activities.

10. The concept of household, on the other hand, is based on arrangements made by persons, individually or in groups, for provision of food and accommodation through pooling of resources such that each member of the household has an equal claim to the resources of the household. The household may be a one-person household or a multi-person household.

11. The population census uses other enumeration units, such as the person, the living quarters and the building. However, for integrated data collection programmes, the most important primary data collection units are the establishment and the household as these are the units which are also used in systems of integrated economic accounts. To the extent possible, concepts and definitions should be standardized. In all surveys there are certain frequent classifiers which are used

such as age, sex, religion, race, ethnicity, marital status, education, activity status, etc. To be able to capture data on these in a consistent manner it is necessary that the same concepts and definitions and questions in questionnaires should be used for all rounds or phases of all surveys.

12. Classifications are another tool for integrating statistical data. Several classifications have been introduced for different purposes. Classifications which are currently being used include classifications of commodities, a classification of occupations, the classification of non-financial assets, the classification of all economic activities, the classification of education, the classification of functions of government etc. These classifications are useful in introducing order and meaning to the numerous data collected in official statistics programmes. Organized systems of data classification facilitate data integration and in this regard, international classification systems such as the Standard International Trade Classification (SITC) and the International Standard Industrial Classification of all economic activities (ISIC) have played an important role not only for data integration but also in harmonizing inter-country classifications. Relationships between one classification system and another, where appropriate, are further made possible by cross classifications and use of bridging-tables.

(c) Permanent field operations

13. Sharing of survey personnel and facilities requires well-trained enumerators and supervisors and that adequate facilities such as transport vehicles, computers, printing equipment, etc are in place. Once the facilities

have been acquired and staff trained, their continuing use contributes to the improving the quality of data collected.

(d) Survey design

14. The costs of frame construction and sample selection for a chosen survey design can be quite high in certain situations. As a result, it might be beneficial and less costly to use survey designs that make use of same frame as was previously developed. Through this approach costs are reduced or spread over several years. This will enable linkage between a household survey and other statistical activities such as population censuses and surveys, economic censuses and surveys, administrative records, etc.

(e) The System of National Accounts (SNA) as an integrating framework for statistics

15. The SNA has played a central role of coordination as a framework for data collection, compilation and analysis. Its conceptual framework ensures the consistency of the definitions and classifications used in different, but related, fields of statistics. Consistency between the different systems enhances the analytical usefulness of all the statistics involved. Harmonization between the 1993 System of National Accounts (1993 SNA) and other major systems has been largely successful and has been achieved by making changes to the SNA as well to the other system.

16. Various other international guidelines are revised and issued at about the same time. Because of the active involvement of International Monetary Fund (IMF) in the revision of the 1993 SNA, the balance of payments statistics, government finance statistics and money and banking statistics are now consistent with the 1993 SNA to the fullest extent possible. It is also the case for the industrial statistics, the standards on labour statistics and the distinction between the formal and informal sectors and the agricultural statistics. It should be mentioned that it is the policy of all the various international agencies at the world level to harmonize their respective statistics systems with each other and with the 1993 SNA to the fullest extent possible.

17. The SNA framework has occupied a central position in economic and social statistics as an accounting framework for ensuring the numerical consistency of data drawn from different sources such as industrial statistics, household surveys merchandise trade statistics, VAT returns and other administrative by-products. The development of satellite accounts within the framework of the 1993 SNA is a typical example for the integration of social statistics with economic statistics. The satellite accounts provide a framework for gathering information in a social field with the following characteristics:

i) it groups in the same structure the presentation of monetary and non-monetary data (physical indicators);

ii) it links the detailed analysis of the area under consideration to the overall economic analysis done by the SNA.

18. Man's activities in exploiting his environment and the resultant concerns for sustainable development have similarly led to the development of a system of Integrated Environmental and Economic Accounting (IEESA) which is another satellite account of the 1993 SNA. The IEESA describes the relationships between the natural environment and the economy and provides a link between economic accounts and environmental and natural accounts.

III. Specific issues related to data integration

(a) Household statistics

19. Household surveys are not a new subject to statistical offices in African countries. Their existence dates back to the time of establishment of the statistical offices. Initially household surveys were launched in order to enable countries to respond to immediate data needs, development planning and formulation of policies and programmes. Agricultural surveys were common in many countries. So also were household income and expenditure surveys which enabled the setting up of cost of living indices and the compilation of national accounts data on private consumption expenditure.

20. During the 1980s a number of African countries conducted their household surveys under the umbrella of the National Household Survey Capability Programme (NHSCP) or the African Household Survey Capability Programme (AHSCP) of the United Nations. The aim of these programmes was to assist interested countries to develop enduring national instruments and skills for survey taking. These surveys also provided a basis for integrated multi-purpose surveys. Other programmes such as the World Bank Living Standards Measurement Study (LSMS) and Structural Dimensions of Adjustment (SDA),

Demographic and Health Surveys (DHS) also promoted the conduct of a number of limited purpose household surveys in African countries.

21. Issues which countries face in the implementation or in planning integrated household survey programmes include the following:

Choice of appropriate design. The choice of a survey design is unlimited. The primary objective is to develop a detailed multi-year plan of household surveys. If there are household surveys being undertaken, the plan should take those surveys into account. If, on the other hand, there are no household surveys being undertaken, it is proposed that a good strategy is to start with a continuing or periodic multi-subject survey and other surveys can be added to the programme later as needs and resources dictate. Any plan development should be guided by three objectives: efficiency, quality and timeliness of results.

Choice of suitable frames. A frame is a list of units such as maps, housing units, census enumeration areas, households, etc. The frame is critical for all stages of sample selection in an integrated household survey programme. Costs of developing frames are quite high, it is therefore useful to use existing frames which have proved acceptable for survey operations. Normally most frames are built for the selection of first stage units and for subsequent stages frames are organized in the field before selection.

Field staff and facilities. These are matters of great concern for the smooth operations of activities in the field. The staff could be permanent or temporary field staff depending on the situation in each country. Whatever is the case, the training and re-training of enumerators is critical for the success and improvement of data quality. Supervisors are usually assigned from the permanent staff of the office and are always at a senior level with many years of experience in statistical operations. Facilities such as vehicles, computers, printing equipment, etc. if lacking or inadequate could contribute to poor quality of survey results. It is therefore essential that all these elements be examined before the survey is launched.

Use of master samples. Many countries have tended to use master samples of one kind or another for their integrated household surveys. A master sample is defined as "a sample from which sub-samples can be selected to serve the needs of more than one survey or survey rounds" (UN, 1986). The great advantage of a master sample is efficiency in the reduction of overall cost of providing samples for multiple surveys or survey rounds. In addition master samples can be cost effective in avoiding duplication in compiling the necessary sampling materials and selecting the samples. Other potential benefits are improvements in the quality of survey results and greater flexibility to respond quickly to needs for data on a variety of topics.

(b) Business statistics

22. As indicated earlier, the term "integration" is used in Statistics to refer to the ability to combine data from a number of different sources in order to provide an

internally consistent account of a given situation. An integrated system of business statistics is therefore regarded as an essential tool not only for development planning, monitoring and evaluation, but also for management of national economies. Its main function is to generate timely, reliable and comprehensive data on all enterprises and establishments engaged in economic activity within the territorial boundaries of a country. The setting up of such a system at the national level implies that all business surveys, including those that are conducted without the direct involvement of the National Statistical Office, should be integrated into a central framework. This requires specific and strict measures, which include:

(i) The establishment of a central body responsible for the coordination of all business surveys; this central body could be the Unit of the National Statistical Office in charge of business surveys;

(ii) The development of a common register of enterprises and establishments, which can be used for all business statistics and which enables the results of different inquiries to be related to each other and form an integrated system of information:

(iii) The harmonization of basic concepts, definitions and classifications used for business surveys; the framework for this exercise could be either the national accounts system or any other system for the assessment of national resources and their uses;

(iv) The standardization of survey reference periods (the calendar year should, preferably, be adopted for annual surveys);

(v) The systematic processing and dissemination of business surveys results in order to provide potential users with valuable information on the economic performance of the country; and;

(vi) The centralization of business surveys results in a database accessible to all users.

23. In principle, an integrated programme of business statistics is implemented in two phases, namely:

i. The organization and conduct of a nation-wide census of economic activities;

ii. The launching of annual, quarterly and monthly surveys programmes.

24. The nation-wide census of economic activities is a key element in an integrated programme of business statistics. It should be followed, as closely as possible, by annual sample surveys that provide a continuous measure of economic activity. Quarterly inquiries on basic indicators, such as employment and output, are the other vital elements in an integrated programme of business statistics. As regards monthly surveys, they should cover industrial establishments only and the

data collected should be used exclusively to compile an index number of industrial production.

25. The nation-wide census of economic activities should consist of the following three phases:

(i) A preliminary full-coverage canvassing of recognizable production units, which will be used to compile a comprehensive register of enterprises and establishments;

(ii) A census (complete enumeration) of large enterprises and establishments; and;

(iii) A sample survey of the small units.

26. The following is a suggested timetable for the initiation of a decennial integrated programme of business statistics :

Year	Type of survey		
	Census	Annual	Quarterly or (monthly)
First year	A		
Second year [a]			
Third year		B	B
Fourth year		B	B
Fifth year		C	B
Sixth year [b]		C	C
Seventh to ninth year		D	C
Tenth year	A	[c]	C

A - All establishments
B - Cut-off sample of large establishments only
C - Probability sample (large sample) – yielding sectoral data for the country as a whole and for most of the principal geographical areas
D - Probability sample (larger than C) – yielding sectoral data for all of the principal geographical areas

[a] No survey is proposed for the second year, which would be devoted to prepartory operations
[b] Countries may choose to conduct a benchmark census during the sixth year
[c] In the census year, the annual survey is integrated into the benchmark inquiry

IV. Institutional arrangements for data integration

27. Coordination of statistical activities embraces integration of data in a national statistical service. Coordination can usefully be included in the statistical law through the establishment of a statistical coordinating body such as the the Statistical Council/Committee. This is particularly essential in a decentralised statistical service. The existence of legal authority may be necessary but not sufficient for coordination to take place. The coordinating body once in place can provide strong leadership and secure the full cooperation of the various egencies involved in statistical activities.

28. A recent survey of Statistical Development of African countries which was conducted by ECA showed that 43.5 % of the countries in Africa have put in place national statistical councils, 30.4% of the countries have producer/producer committees and 43.5% of the countries have in place user/producer committees. These give an indication of attempts by African countries to foster data integration through committee structures. There is however need to do more in this area to ensure that statistical activities are adequately coordinated and data integration is promoted actively.

29. User/producer and producer/producer committees can be internal (within the statistical service) or external (with other producers/users of statistics within and outside the statistical service). In the case of user/producer committees, they could be used to access needs and coordinate activities relating to various statistical activities particulaly at the sectoral level. Producer/producer committees can also be very

effective tools for discussing the various integration tools: concepts and definitions, questionnaires and their control, registers of statistical units, statistical standards, etc.

V. Conclusion and recommendations

30. The quality of official statistics can be increased if the data collection programmes are addressed holistically and all interrelationships are fully exploited. National statistical services should, whenever possible implement data collection programmes which utilize the advantages to be gained from data integration. Those advantages include increased data reliability, enhanced data coherence, and optimized use of limited resources. Household surveys, therefore should be organized in an integrated manner so as to meet multi-purpose data requirements.

31. For business statistics, it is recommended that African countries should establish a decennial timetable for the implementation of an integrated system comprising economic censuses, annual, quarterly and monthly surveys as described above.

References

1. Statistical Papers, Series M, No. 67/Rev. 1 (United Nations Publication, ST/ESA/STAT/SER.M/67/Rev. 1, 1998.

2. National Household Survey Capability Programme (United Nations Publication, 1986 DP/UN/INT-84-014/5E