Statistical Journal of the United Nations ECE 1 (1983) 275–283 North-Holland Publishing Company

# Editorial

The present issue is devoted to managerial problems facing the executive heads of national statistical offices. Its contents are largely based on the contributions submitted to the 'Seminar on statistical data collection and processing systems under new conditions' (21–25 September 1981). The Seminar was convened by the Conference of European Statisticians and held in Moscow at the invitation of the Government of the Union of Soviet Socialist Republics.

The Seminar was intended for officials concerned with statistical policy issues at the highest level. Its purpose was to discuss problems relating to the development of statistical data collection and processing systems in view of the rapidly developing computer technology. The Seminar addressed four problem areas.

(1) The interrelationship between national statistical offices and other organizations concerned with the collection, processing and analysis of statistical information.

The article prepared by Malinvaud was submitted to the Seminar under this item in its agenda.

The Seminar engaged in a vivid discussion on the basis of the papers submitted. The discussion focussed on the role of statistical services in the overall national information system as modified by the spreading of computer technologies. Such technologies enhance the establishment of an integrated national information system. Within this broad perspective the Seminar considered (i) the specific contribution of statistical services to the promotion of an integrated information system; (ii) the main individual problem areas to be tackled by statistical services while making their contribution; and (iii) the repercussions of an emerging integrated information system on the internal organization of statistical services as well as the future role of statistical surveys.

National statistical offices play an essential role in providing data to the highest decision-makers in Government, and to organizations, institutions, enterprises, as well as individuals. Statistical information is collected mainly through censuses, regular sample surveys and ad hoc inquiries. However, the rapidly increasing demand for data at all governmental levels, in combination with new conditions now prevailing in the processing of information have enhanced the possibilities to use administrative and other sources of data for statistical purposes.

The rapidly increasing number of suppliers of administrative and statistical data outside national statistical services offers new sources of data; hence the new potentialities for making the available information more complete and useful. This not only has considerable effect on the structure, organization and function-

ing of national statistical agencies, but also leads to a series of problems of a technical nature requiring new approaches to their solution.

Ad hoc surveys are usually carried out independently by a number of government departments; administrative files containing information about various sectors of the economy are regularly maintained by tax authorities, social security and other specialized agencies; private organizations operate in the field of social and economic research, dealing with specialized subjects. In this situation the interrelationship between national statistical offices and other organizations active in the communication, collection, processing, analysis and dissemination of data needs to be clearly defined. The organization of the decision-making processes has an important impact on the nature of this interrelationship. For example, the elaboration and implementation of national plans for economic and social development require a continuous provision of information at all levels of state control - a situation which has consequences for the organization of the national statistical service, including the structuring of the information flow between statistical offices and other administrations. In general, the relationship between statistical offices and other bodies, which also collect and process economic and social data, calls for some co-ordination. The co-ordinating role of a national office of statistics is of primary importance in the following respects: (i) the technical co-operation between the country's statistical bodies, ministries, government departments, and centers dealing with data processing; (ii) the control and evaluation of the basic information as well as of the statistical aggregates derived from it; (iii) the forecasts and anticipation of new information needs; (iv) the identification of suppliers and final users of statistical information; (v) the determination of priorities for the satisfaction of different information needs; and (vi) the design and specification of administrative data. A particular aspect of co-ordination is to be found in the integration of statistical data with various administrative and other economic and social data. The degree to which such statistical integration can be easily achieved depends partly on the extent of centralization prevailing in the organization of the statistical system.

It was stressed that statistical offices may and should play an essential role in developing and implementing an integrated approach to the collection, processing, analysis and dissemination of data. The prevailing level of centralization of statistical services was identified as a major element in the implementation of an integrated approach. Another major element was identified in the possibilities open to statistical agencies to influence and perhaps modify the contents of administrative files. Several participants reported on their capabilities in this respect. The task was generally recognized as being difficult. The increasing dependence of statistics on external data sources may lead to a re-evaluation of traditional quality criteria as well as classification principles.

The process of data integration is accelerated by using modern electronic data processing techniques. Automated methodological data bases, data communication networks and other advanced computer facilities can promote such integration, in particular, when the structure of data banks and their organization are geared towards this objective. The co-ordination of statistical activities or, more

specifically, the co-ordination of statistical data files through their integration, the automation of statistical work and the rapidly increasing propensity to handle administrative data by automated management can be achieved only if methodological and technical compatibility can be ensured. The required compatibility relates to both the statistical system and the joint functioning of the statistical system with other automated systems.

Methodological compatibility is ensured first of all by unifying the systems of indicators, classifiers, registers, concepts and nomenclatures used. In addition, meta-information systems have to be developed, which include data dictionaries, general catalogues, documentation of numerical data with indication of sources, definitions and comments on methodology. The methodological compatibility is essential to facilitating direct access to data, especially in an interactive environment, and to preventing any erroneous interpretation of data. Technical compatibility is achieved by the use of technical facilities for the collection, processing and transmission of information.

Special arrangements have to be made for preserving confidentiality of data in view of the access to administrative information by statisticians and the increasing computerization of statistics. The Seminar emphasized that confidentiality of individual personal data has to be strictly observed. This can be achieved by making available to the public only aggregate results when matching different administrative files for statistical purposes and only material relevant for general studies when data are gathered from sample surveys.

The Seminar expressed its general agreement on the fact that statistical surveys will always be needed. Firstly, not all required data are recorded in administrative files. Secondly, the statistical quality of administrative information is sometimes doubtful; administrative data do not always pass through consistency and comparability checks necessary for their statistical use. Thirdly, it may be easier to derive various types of information from one special sample survey than from matching different files. Different files do not always cover identical units of observation nor are they necessarily based on identical definitions. Thus, their matching may be particularly difficult if time series have to be constructed from them – always assuming that no problems of confidentiality arise.

The Seminar concluded discussion of this topic by stressing the leading role of central statistical offices in structuring all kinds of statistical information available within countries and in disseminating it to users. It also stressed the importance of developing close co-operation and communication between statistical and other public administrations at all suitable levels.

(2) The role of modern techniques in the regionalization of statistical activities. The Seminar discussed this item on the basis of four contributions, among which was the article by Fastbom. The term 'regionalization' is subject to interpretation. Consequently, the Seminar first of all clarified the meaning of the term. At a second stage, the major determinants for the optimal level of centralization of statistical services were identified. Thirdly, the scope of possible decentralization with regard to individual functions of statistical services was explored. Finally, the

Seminar considered some of the central tools and difficulties in efforts to increase decentralization.

Several meanings of the term 'regionalization' were identified. Firstly, regionalization can relate to the organizational structure of national statistical systems. Secondly, regionalization could occur in relation to functions to be performed by the overall system of national information, whether inside or outside the national statistical agency. Thirdly, the term could reflect a dispersion of technical facilities over a country without a major impact on the organization of the national statistical system. These various interpretations are neither mutually exclusive nor exhaustive.

Several participants reported about the presently existing structure of the national statistical system in their countries. It was agreed, however, that such structures could not be taken as fixed. Under present circumstances prevailing in many countries, the extent to which a national statistical system should be decentralized has to be viewed as the result of a dynamic process which is continuously influenced by many factors. The size of a country, the overall organizational characteristics adhered to, as well as the detailed administrative practices followed, were identified as being among the major determinants of the degree of statistical centralization. The recent as well as the expected advance in computer technology could mainly act as an element removing important constraints from the organization of the activities of national statistical offices in respect of centralization–regionalization. In this sense the new technological conditions may bring about qualitative changes in the organization of information flows.

A general trend towards an increased relative importance of decision-making at the local and regional as opposed to the national level is observed in several countries. The request for more statistical information of a purely local or regional significance accompanies this general trend. The ensuing information needs can either be satisfied through the administrations themselves in which the need is generated, or through the statistical service, or through a mixture of both. The adaptation of the organization of the national statistical service to such circumstances would have to maintain the unified classifications and methodological standards of data at whatever level they are collected. It was noted that the initial position from which the necessary adaptation is attempted either complicates or facilitates its achievement. In centrally planned economies, the collection and production of statistics is normally organized in a hierarchical manner, which guarantees the consistency of statistical reporting at each administrative level. In market economies, where there is generally a multitude of agencies collecting and producing data, considerable efort is required to overcome the danger of disseminating statistics which cannot be easily compared.

Concerning the decentralization of the functions performed by the statistical system the Seminar made a distinction between the collection of data, the establishment of a micro-data base, the aggregation of micro-data, their dissemination and their analysis. Several participants argued in favour of a regionalization of the collection of data, particularly with regard to household surveys. The

Seminar was informed of the experience gained in several countries having implemented regionalized data collection. Such experience is the more positive the more important the reluctance on the side of respondents is to report to centrally managed requests for information and the better equipped the national statistical system is to ascertain the consistency of data collected in different regions of a country.

The problem of establishing a nationwide and consistent micro-data base can only be solved with reference to the availability of suitable computer hardware and software. If the central statistical office is capable of ensuring homogenous processing of collected data, a decentralized approach may be as effective as a centralized approach and even preferable on budgetary grounds.

The aggregation of micro-data should primarily be done in a centralized manner. However, aggregates needed for local or regional decision-making could also be obtained in a decentralized way. This solution may particularly be advantageous in circumstances in which regional statistical offices perform the function of analyzing the social and economic development of their region in close co-operation with the responsible levels of general administration.

A similar consideration applies to the organization of the dissemination of statistical information. The predominant concerns relate to the provision of information which is not only adapted to the needs of specific users but also comparable at both the regional and national levels. The satisfaction of all specific user needs at the local and regional levels may be left to regional agencies, but the central office would have an interest in the maintenance of certain features permitting a wider comparability of disseminated data.

The Seminar emphasized the importance of registers as a tool for the general integration and co-ordination of statistical systems. Hence, such registers could also have to play a role in any regionalization of statistical activities. The management of such registers may be centralized or decentralized, but it is important to develop and maintain consistency between registers.

Several participants reported about specific problems encountered in their countries during efforts of decentralization, as well as on some results thereof. It became apparent that the breakdown of important general statistics such as census results by region normally does not pose major problems. On the other hand, serious difficulties are often encountered in integrating sample surveys as well as administrative files for different regions. If these problems can be solved, regionalization of statistical services may improve the timeliness of data, as well as other quality attributes of statistics.

In conclusion, the Seminar stressed the necessity of preserving a unified centralized methodology as well as of using common software. It was agreed that the new technology removes constraints from communication. Hence, the possibilities for statistical offices to decentralize their activities are increased. At the same time, the basic decision on the extent of centralization has to be taken with regard to the optimal adaptation of statistical services to the general circumstances in which these services are rendered. (3) Methods used to ensure the quality, reliability, timeliness and confidentially of statistical information.

The articles by Salapa and de Vries were among the papers submitted to the Seminar under this item.

The discussion papers submitted to the Seminar for consideration enumerated several quality attributes of statistics. The Seminar discussed the most important attributes in relation to the introduction of computer technology into statistical services under two aspects. On one hand, the maintenance of the traditional quality attributes requires new approaches in a computer environment; on the other hand, new quality attributes emerge through the introduction of automated data processing.

In accordance with the rapidly changing EDP technology and related EDP environments statistical services have passed through different phases of automation and may now be planning or executing new or modified strategies. This development affects directly the quality attributes of statistics. More specifically, statistical quality is normally defined and/or measured on the basis of certain attributes such as statistical accuracy, relevance, reliability, compatibility, timeliness, flexibility and efficiency. These attributes are reflected in the information system, which in turn has repercussions on the attributes. Thus, the overall effectiveness of the information system depends, among other things, on such quality attributes. Hence it was suggested that a measure be developed which could give insight into this relationship. It was recognized that the development of a unique measure would pose many problems. Alternatively, the joint use of several quality indicators could be considered.

Three broad categories of problems can be identified in setting up and operating a statistical production system. Their solution implies a particular combination of quality attributes. The questions dealing with statistical co-ordination fall into the first problem category; they include basic requirements which are essentially of a statistical nature. The second problem category covers questions which arise when using new techniques or methods of data processing; they are essentially EDP problems. Last but not least appear problems with regard to organizational and managerial aspects associated with the introduction and implementation of a particular information system.

Essential elements of statistical co-ordination are data compatibility and integration of individual statistics into a coherent system of statistics. Co-ordination of statistical activities can primarily be achieved through the setting up of central registers from which statistical units are selected on the basis of their recorded characteristics. Such registers are of great help in controlling the completeness of statistical information and improving the accuracy of data. They should be used to a large extent even if considerable costs are involved. The units contained in the registers can be classified according to available statistical classifications or nomenclatures. The co-ordinating role of statisticians consists essentially of ensuring the uniformity and accuracy of definitions and concepts used in statistics; this will facilitate the circulation of economic and social information which consequently becomes more useful and available to a wide public.

The use of advanced electronic data processing techniques, the technical standardization of equipment and the use of modern hardware environments have important effects on almost all statistical quality attributes as well as on the volume of statistical output. Increasing user requirements can be met through these developments. The positive experience gained by various countries in using mini-computers was stressed. It can be expected that this new and rapidly developing technology will play an important role already in the near future. Timeliness, flexibility and productivity may profit from the introduction of the methods mentioned above. On the other hand, standardization of equipment and generalized software may cause some rigidity in the presentation of statistics to users, i.e. in output formats.

Comparability of data and data integration are required for several reasons, among which is the efficient utilization of new techniques. The production process will greatly be affected by using data base techniques which make information systems available to a larger number of users. Access to and dissemination of information are facilitated if the system operates in an interactive environment. Timeliness, harmony and flexibility seem to be quality attributes which are positively affected by this methodology. Improvements will also lead to a more efficient overall statistical system.

The maintenance of confidentiality of individual information constitutes a technical problem if data bases, integrated in a common information system, can be directly accessed by a wide public. There was agreement on the proposition that 'statistical privacy' should always be guaranteed. Technically it is possible to preserve confidentiality by ensuring that only authorized users have full access to data, or by making available only such micro, macro and aggregate data, from which nominal identification has been removed.

A new information system also has an impact on organizational aspects of the statistical service: restructuring of functions, decentralization or internal coordination may be required to create the conditions in which the system can be operated more efficiently. Shortage of resources, increasing requirements for professional staff with particular types of training, and cost allocation also have to be taken into account in the establishment of a new production model.

Attention was drawn to the specific problems which are encountered in this context by international statistical agencies. Such problems relate mainly to the publication of internationally comparable data of adequate quality. The collection, processing and compilation of national data should not only minimize the response burden on countries, but should also facilitate the provision of high quality international statistics.

In conclusion of the discussion under this topic it was agreed that the production process of statistics should be optimized with respect to quality attributes. The development of new technology in data processing may give rise to new quality attributes such as the development of an integrated and co-ordinated data base of sufficient flexibility.

(4) Experience gained in the training of staff in the use of EDP technology for the processing of statistical information.

Below contributed to the documentation discussed at the Seminar in the context of this topic. This discussion focused on the problem of how to bridge the gap in communication between statisticians and EDP experts. The gap is apparent in many countries. Three basic options for solving this problem were identified.

Firstly, statisticians could be trained in the use of EDP technology, particularly programming, for the purpose of guiding EDP experts more effectively. Secondly, EDP experts could be trained in practical statistical methods so as to increase their understanding of the use which is made of data bases in the production process of statistics. Thirdly, a separate group of staff could be formed that would combine both types of skill in such a way that the necessary communication between statisticians and programmers could be channelled through them.

The Seminar agreed that the basic divisions of labour between statisticians and computer experts would normally have to be maintained. However, the expected future development of computer hardware and software will not only influence the attributes of both types of professional skill, it will also have an impact on the relative weight of various functions within the overall composition of the staff of statistical offices. The qualifications required within the group of computer experts may in the future be concentrated on the development of general programs, co-ordination of data and data management. On the other hand, statisticians may acquire the necessary qualifications to be able to communicate directly with the computer.

Several participants reported about ongoing training programmes and the experience gained therein in their countries as well as the planned direction of future training efforts. Existing programmes mainly aim at enhancing the capability of statisticians to understand the functioning of computer operations. It was recognized that such training programmes vary from country to country, primarily in relation to the training facilities for computer experts which are available in the national educational system outside statistical offices.

In conclusion, the Seminar agreed that the recognition of new skills for the conduct of professional work in a changing environment constitutes a generally valid phenomenon. In the present context, efforts have to be made in order to ensure the availability of both skills required for the adequate functioning of statistical services. At the same time, facilities have to be provided to individual staff in such services which permit staff to adapt to emerging professional needs. The success in both these efforts is of crucial importance to the proper evolution of statistical services.

The full proceedings of the Seminar will be published by the Central Statistical Board of the USSR.

The contribution by Jensen to this issue of the Statistical Journal treats a problem of outstanding importance to statistical offices. In Denmark, as in several other countries, efforts are made to use administrative registers to the maximum extent possible for the purposes of official statistics. Related policies are prompted by a variety of reasons ranging from attempts to reduce the

response burdens in relation to statistical surveys to the objective of making full use of existing computer technology. At the same time, a number of problems are encountered in this approach. The article by Jensen is an amended version of a document presented for substantive discussion to the thirtieth plenary session of the Conference of European Statisticians. The article discusses the relevant experience obtained in Denmark in the recent past.

.