STATISTICAL COMMISSION
Twenty-first session
17-26 February 1981
Item 7 (g) of the provisional agenda

SOCIAL AND DEMOGRAPHIC STATISTICS: LINKS AMONG
ECONOMIC, SOCIAL AND DEMOGRAPHIC STATISTICS

Role of macro-data and micro-data structures in the integration
of demographic, social and economic statistics

Report of the Secretary-General

SUMMARY

The present document responds to the request of the Statistical Commission at its twentieth session for a continuation of the work on the integration of demographic, social and economic statistics and, in particular, for work on the structure of over-all linkages among classifications and the role of the national accounts as an integrating framework. It continues the work reported in The Development of Integrated Data Bases for Social, Economic and Demographic Statistics.

The document discusses the need for an integrated data framework, considers the nature of such a framework and the general methodology for constructing it, reviews the work programme approved in 1979 and suggests priority areas for future work. Points for discussion by the Commission are included (para. 80).

80-16410

/...

CONTENTS

CONTENTS (continued)

INTRODUCTION

1. The present report is a continuation and further development of the path indicated in the recent Statistical Office publications on methodologies for integrating social and demographic statistics, namely, Studies in the Integration of Social Statistics: Technical Report 1/ Improving Social Statistics in Developing Countries: Conceptual Framework and Methods 2/ and The Development of Integrated Data Bases for Social, Economic and Demographic Statistics. 3/ At the same time, however, it will take into account the need for co-ordination of the work on the Framework for Social and Demographic Statistics (FSDS) with the concurrent work on the extension and updating of the System of National Accounts. (SNA).

2. The report will explore further some of the issues raised in earlier discussions. The focus of interest is on the ways in which the different special fields of statistics impinge upon one another. Earlier reports have pointed out the kinds of problems that arise from the failure to consider the interrelationships among different fields and have summarized some examples of efforts to solve the problems. The present report will attempt to synthesize this experience and to draw from it proposals for the organization of multipurpose, reusable demographic, social and economic micro-data bases. Beyond this, it will explore the relation of such micro-data bases to macro-data structures, in the hope of making some progress towards the construction of an over-all framework embracing all levels of information.

3. In addition, the report will suggest where new developments are occurring, on both the macro level and the micro level. On the macro-economic level, attempts are being made both to broaden the scope of the economic accounts through the addition of information on non-market activities, environmental factors and a range of other topics and to improve their usefulness for the analysis of social and distributional questions by the addition of much more extensive disaggregation into social and demographic subsectors. Broadening the scope of the accounts was discussed in the report entitled The Feasibility of Welfare-Oriented Measures to Supplement the National Accounts and Balances: A Technical Report 4/ and is considered further in the report before the Commission under agenda item 4 (a), entitled "Future directions for work on the System of National Accounts (SNA)" (E/CN.3/541). But it is not only for economic statistics that macro-structures are important. There is widespread interest in the development of measures of the level of living that go beyond the elements covered in the national accounts. In another approach, also, some researchers have been experimenting with the construction of demographic accounts, where there is an attempt to introduce system and order into a number of aspects of demographic data and analysis.

1/ United Nations publication, Sales No. E.79.XVII.4.

2/ United Nations publication, Sales No. E.79.XVII.12.

3/ United Nations publication, Sales No. E.79.XVII.14.

4/ United Nations publication, Sales No. E.77.XVII.12.

4.   On the micro level, the National Household Survey Capability Programme will contribute greatly to the availability of micro data. It is important that, as the supply of data and the technical capability to handle it grow, attention be focused on the organization of the data into a coherent over-all framework.

I. NEED FOR AN INTEGRATED SOCIAL AND ECONOMIC DATA FRAMEWORK

5.   Over the course of the last three decades, the Statistical Commission has continually been concerned with the development of statistical frameworks for organizing economic and social data. Its work has included both macro-level data systems such as the System of National Accounts and recommendations at a much more basic level, such as those on censuses (population, housing, industrial) and surveys. Its work on such methodological and systems development has played a very important role in the enormous growth in the availability of organized and standardized statistics on both the national and international levels.

6.   Building upon this foundation, it is now possible to explore the possibility of erecting a more generalized data framework that will not be limited, as work to date largely has been, to either economic data, on the one hand, or social and demographic data, on the other, but will embrace both. This is increasingly necessary, as the kinds of problems facing policy-makers today cannot be neatly categorized into one field or the other. Information is needed that will help, for instance, to assess the impact of economic problems like inflation and economic stagnation upon various social groups and to identify those most vulnerable and, in a period of increasing scarcity of some resources, to discover who suffers and to measure the impact upon the level of living and its distribution. Policy-makers need to know what the mechanisms are: who saves, who invests; who benefits from social programmes, who pays their costs; how the benefits compare with the costs, how they are distributed among social groups; how social change takes place, what impact economic factors have upon it and what the implications are for policy. In all of these questions and many more, social and economic aspects are intimately related, and ways are needed to bring both sorts of data to bear.

7.   The need to bring demographic, social and economic data into a closer relationship is now seldom questioned. In the work on the integration of social statistics, it has been recognized from the outset that one important facet of social statistics is their relation to economic statistics. Of the 11 fields of statistics identified in Towards a System of Social and Demographic Statistics, 5/ for example, 4 are primarily concerned with "economic" topics, and in another 3 economic information occupies an important place. Conversely, extensions of the macro-economic framework to meet new needs inevitably lead directly into social and demographic considerations. It is not enough to know that unemployment or prices have risen, or that the rate of growth of output has fallen; to diagnose the problems properly, it is necessary to know how unemployment is distributed among age and sex groups, ethnic groups and occupations; how the cost of living of different social groups has changed and how the levels of living of the elderly, the urban poor, the farm population, the small entrepreneur have responded; what share of the cost of slow growth is paid by households, by business and by government.

_____

5/ United Nations publication, Sales No. E.74.XVII.8.

8.   In order to achieve the kind of integration of the statistics that is required for an analysis of such problems, it is necessary to take into consideration all of the available statistics, and it must be possible to make use of them at whatever level of aggregation a particular problem calls for. Social and economic data come from a wide variety of sources: complete census enumerations, sample surveys designed to be comprehensive in coverage, sample surveys addressed to particular questions of varying scope and a great variety of administrative sources. Most of the macro structures that have been constructed in the past, whether economic (such as national accounts) or social (such as demographic accounts), have made use of a variety of sources and have endeavoured to bring them all into a consistent picture.

9.   But consistency and coherence on this level, although essential pre-requisites are no longer enough. Today's technology makes it feasible to ask for data that are coherent throughout, from the lowest to the highest levels of aggregation. Thus, users interested in the impact of over-all fiscal and monetary policies should be able to examine not only the aggregate effects on gross domestic product as a whole and the major national accounting components, but also on specific social and demographic groups. Conversely, those interested in a particular social programme - such as aid to disadvantaged children - should be able to examine the costs and benefits of the programme in the context of the lives of the client group of children, in the context of all children, in the context of the government budget and in the context of the over-all state of the society and economy. Such an ability to move back and forth among the levels of aggregation and between related fields is essential if problems are to be seen in an appropriate perspective.

10. This ability is also essential, however, for another and perhaps more important reason. Broad aggregates show what is happening to the society or economy as a whole, but they cannot show the ingredients that go into a given aggregate result. An aggregate may change either becuase the behaviour of the individual units of which it is composed has changed or because the mix of individual units has changed. A fall in the birth rate, for instance, may reflect either a fall in the birth rates of both rural and urban women, or merely an increase in the proportion of women who are urban, without any change in the behaviour of either group. The two situations would look the same in the aggregate statistics, but they are in fact quite different and would call for quite different policy responses. In the absence of integrated data, statisticians and econometricians have devoted much mathematics and many pages in academic journals to discussion of "the aggregation problem", with a notable lack of productive results. There is still a widespread tendency for analysts to assume that aggregate behaviour is micro behaviour written large - an assumption that often leads to inappropriate policy choices.

11. To achieve the required level of integration, the starting place must be integration of the information that is available at the individual or household level. It is by now generally recognized that, from a purely technical point of view, the most efficient way to preserve data for future use is in the form of micro data; any aggregation reduces the potentiality of the data for future unanticipated uses. What needs further exploration, however, is the question of

/...

relating various bodies of micro data to one another. Furthermore, not all relevant social and economic data are available, even conceptually, at the individual or household level. Other kinds of reporting units are often relevant - the village, the school, the government ministry, even the establishment or enterprise - and ways must be found to bring these kinds of data into the information framework.

12. The same principles of organization apply throughout the social and economic fields. They have guided international work on systems and standards from its inception, but they have been particularly emphasized as the growth of technological (largely computing) capability has expanded the realm of what is feasible. They were clearly brought out in the study Sir Claus Moser prepared for the Statistical Commission at its eighteenth session (E/CN.3/449 and Corr.1), in terms of consistency, comprehensiveness and linking; later work explored in a preliminary way the implications of these principles for the construction of integrated data bases.

13. Section II below will examine the nature and requirements of a framework for demographic, social and economic statistics. Section III will discuss the contribution that the United Nations Statistical Office can make to the practical development of work in this area.

II. DEVELOPMENT OF THE FRAMEWORK AND METHODS OF INTEGRATION

A. Nature of the framework

1. Framework for social and demographic statistics and its potential

14. The Framework for Social and Demographic Statistics embodies the work done to date at the international level on the construction of an integrating framework for social and demographic statistics and, as noted above, it also includes a great deal of economic information. The Framework for Social and Demographic Statistics has been the focus of international attention for over a decade now, and a substantial amount of progress has been made. Much of the work has concerned the identification of the subject-matter fields that such a framework ought to cover. But attention has also been devoted to clarification of the nature of the framework itself, and it is primarily the latter question that is of concern here. It has proved to be difficult to arrive at a common understanding on this score, but a consensus is gradually emerging.

15. The Framework for Social and Demographic Statistics should be thought of literally as a framework - the bones and muscles of the statistical system, which both provide structure and support and hold the body together. What distinguishes a framework from a disjointed list of statistical series is structure and coherence: it all has to fit together and to retain its shape and form. Such a framework requires, in the first place, the use of consistent classifications and definitions throughout the entire body of statistics. In the second place, it requires the development of suitable aggregates: there must be a system of aggregate data to which the more detailed and disaggregated data can be related.

16. It is the multidimensionality of demographic, social and economic statistics that makes the specification of the framework an intimidating task. Early approaches to the problem took the route of trying to specify in advance a limited list of time series, drawn from the almost infinite range of possibilities, that might be expected to meet most needs for social data. This followed the traditional practice of the pre-computer period, when statistics were of necessity looked upon as pre-specified cross-tabulations, which had to be reducible to the printed page. This approach ran into difficulties, however, because it proved to be impossible to establish criteria for selection of series to be included. The needs for data are far too varied, and they change as policy and analytic concerns change. A framework constructed in this way runs the risk of early fossilization; it may be preserved as a relic, but it will not often be used, since most analytic questions are likely to require data that are not included.

2. Outlines of a general framework

17. It has become increasingly clear that what is needed is a framework that can accommodate whatever data exist and will find a place for new data as they appear. Despite the diversity of social and demographic data, it is possible to impose some order. As a preliminary approach, consider a universe of r reporting units,

each of which may contain s subunits. Each reporting unit may have k attributes and each subunit j further attributes. Each j and k may take a range of values, some continuous, some not. The problem is to construct a framework that will preserve the data about the reporting units and their subunits in such a way as to make it possible to draw off aggregates, or distributions at any level of detail, of any of the k or j attributes, alone or in combination, all consistent with one another. The framework should provide a structure to which any and all data can be attached, as long as the data relate to the r reporting units (or some of them) or their s subunits. For a different universe of reporting units, there will be a different framework; however, if the two universes intersect, share attributes or impinge upon one another in other ways, they will need to be interlocked through common attribute values and common aggregates of attributes.

18. The construction of such a general framework is still a long way off, but its character is gradually becoming clearer. It can, furthermore, be approached by gradual steps. Unlike such formal structures as input-output tables, for instance, it is not necessary to begin with the ultimate end-product. It is undoubtedly too early to think in terms of a single data structure that will embrace the entire range of demographic, social and economic data. But it is not too early to think in terms of sets of interlocked data structures, each of which would be useful by itself, which together would form an integrated information system. Each such structure would deal with a limited range of topics, and most would rely primarily on single data sources; but through the use of common definitions and classifications and overlapping attributes, it would be possible to move easily from one to another. From the FSDS point of view, the logical place to start is with households. Section B below considers the problems of constructing such a data structure for households.

B. Methodology of linking household data

1. Techniques and tests of validity

19. What is needed is a framework, initially, for data about households and the families, extended families, sub-families, tax units, spending units and individuals to which they are related, with ways for keeping all of the interrelationships straight. What is being sought, in fact, is a mirror image of the actual population, reflecting all of the attributes about which information is available -- that is, a model population. Such a model population can be based on an appropriate sample drawn from the population census. The desirability of establishing such a census micro-data base is now well accepted, as are the principles for constructing it. The question is discussed in some detail in the Principles and Recommendations for Population and Housing Censuses 6/, where the creation of such a micro-data base is recommended.

6/ Statistical Papers, Series M, No. 67 (United Nations publication), in press.

20. But the model should not be static in time or limited in coverage to what can be collected in a single population census. To some extent, it would be possible to extend it over time by exact matching with earlier and later censuses, and its topical coverage could also be expanded by exact matching with other appropriate data bases (usually administrative, on such topics as health, education, social security or taxes) if the data bases contained the same reporting units and considerations of confidentiality did not prohibit this procedure. But exact matching is at best a limited tool, and it cannot be expected to yield the required over-all framework.

21. What is needed is a way to provide a place for everything that is known, that is, to map the information from diverse sources onto the model population, just as a cartographer brings together information from many ground surveys, aerial surveys and even satellite photographs in making a map of the geographical terrain. In order to do this, it is imperative to recognize that the model households, families and individuals in the sample are just that - models - not actual people. Each individual unit in the model population represents the class having a particular collection of attributes. This feature of the model has the useful property of substantially alleviating the problem of confidentiality; since the units in the model are not real people, no one's privacy is being invaded.

22. In constructing the model, advantage can be taken of the observed distributions of a number of attributes to estimate the distributions of - that is, to map - a new attribute obtained from a new source. If, for example, the basic census sample which is being used as the model frame has information on each individual contained in it giving age, sex, occupational status and place of residence, and another (smaller and less comprehensive) sample is also available that contains all of this information plus educational attainment, educational attainment can be mapped onto the model population in ways that will reproduce the distribution of educational attainments associated with any particular combination of the other attributes. A number of techniques have been developed for accomplishing this.

23. The objective, to repeat, is not to reproduce real individuals. It is to produce a model population whose attributes, when aggregated, cross-tabulated, regressed or otherwise manipulated, produce the same results as are obtained from known data. Nor is the objective to manufacture new information. It is simply to bring out and make conveniently available for further use the relationships already existing in the data. In order to do this, it is necessary to provide a repository for all the information that does in fact exist, that is, to attach it to the framework. To return to the previous example, suppose that the census file contains information on an additional attribute - ethnic origin, for example. Since this attribute does not exist in the second smaller data set, it cannot be used as a factor in the mapping; within any age-sex-occupation-region cell, the conditional joint distribution between educational attainment and ethnic origin must be considered to be random, unless there exists a third source from which information on this point can be obtained.

24. Nevertheless, it may happen that both ethnic origin and educational attainment are closely correlated with the attributes that are present in both samples; in

that case, the raw joint distribution between ethnic origin and educational attainment that emerges will be a true reflection of the actual population. A tribal group, for instance, may be predominantly rural and agricultural, and predominantly of low educational status. The mapping would show this. Of course, the mapping shows results, not causes. Continuing the previous example, the model may show that professional people -- doctors and lawyers - are likely to be highly educated and more likely to be of one ethnic group than another. The actual causality in this case is likely to run from education to professional status, and perhaps from ethnicity to education, but not from ethnicity to professional status directly, and no conclusions can be drawn on this point without independent investigation of this specific question. The map shows what any map shows: the features of the terrain. It does not show how they were created. Even though one cannot say why it is so, it is possible to observe that there are few highly educated people of certain ethnic backgrounds and few professional people who are not highly educated, and this may be relevant in policy formulation.

2.   Sources of data and control totals

25. The basic test of the validity of the model is its ability to reproduce known relationships. That is the fundamental criterion behind all of the mapping techniques currently in use. Cross-tabulations should sum to the correct marginals, and regressions should yield the correct coefficients. Where nothing is known about a particular relationship, the model can of course do no better than reproduce the lack of information, by assuming that the relationship is random. But it is possible to, and the model should, integrate all known relationships.

26. What is known will, of course, come from a wide variety of sources. The population census and continuing household sample surveys play a fundamental role in establishing the frame and setting the control totals relating to the society as a whole. But many other sources of data make important contributions to the fund of knowledge about the state of the society and the direction of its change. Often, these are administrative sources: tax authorities, school systems, the social security fund administrators, health care services, the criminal justice system. It is to be expected that conflicts will arise among data sources (or even within the the same source over time). Such conflicts are well known to those who deal in the construction of macro models: obtaining closure of the demographic accounts or reducing the statistical discrepancy in the GDP account are familiar problems. But this, essentially, is the object of the exercise. Bringing the data from different sources together reveals where the problems are and suggests paths for improvement.

27. Control totals -- aggregate data systems - play a very important role in the framework. There are many dimensions that such aggregate constructs can reflect. The basic demographic accounts showing the population by age and sex and its change over time bear an obvious relation to a micro-analytic model based upon a sample drawn from the population census, and both must be part of the framework. Towards a System of Social and Demographic Statistics and later work by other scholars have suggested other dimensions for which similar aggregate accounts can usefully be contructed, including education, migration, marriage, employment and occupation.

28.   One of the most important sets of control totals to which the model must be calibrated is the data contained in the household sector of the national accounts. From the social point of view, income and expenditure, labour force participation and government benefits received are important (and perhaps the most important) ingredients in levels of living and, by extension, in social welfare. From the economic point of view, social and demographic characteristics are important determinants of economic behaviour, and essential to its understanding. The calibration of the social and demographic data to the household sector of the national accounts is the key to integration of social and demographic data with economic data. To serve this purpose well, however, the national accounts themselves may need to be re-examined with this in mind. Some of the kinds of questions that arise are discussed in "Future directions for work on the System of National Accounts (SNA)", (E/CN.3/541), which is also before the Commission.

C.   Linkages of household data to other bodies of data

1.   Governments

29.   As noted in paragraph 11 above, the household (or individual) is not always the focus of social and demographic data, and it is often conceptually inappropriate to arrange the data using the household or any variant of it as the reporting or reference unit. Much social and demographic information reflects the activities of governmental units and is most conveniently arranged using these as the reporting unit. For government programmes, the problem of confidentiality is in some respects more manageable than it is for household data. Budgets and statistics on services rendered, at least in the field of social services, are usually public information. But this is true only of relatively aggregated statistics. Information about small areas and information about individual programme participants may still involve confidential information, so that the same modeling technique suggested above with respect to households may also be useful for governmental activities. One may wish, for example, to include only a sample of local governments where such governments are very numerous, a sample of school systems or a sample of hospitals. One may also wish, however, to include a sample of the households that are affected by the activities of these governmental units, either as taxpayers or as recipients of government services and benefits. It is here that the government model will impinge upon the household model. The attributes of the taxpayers and benefit recipients identified in the government model must be consistent, in definition and classification, with the attributes of the members of the model household population, and relevant control totals must agree.

30.   Consistency in definition and classification of attributes, however, is by no means the same as inclusion of the same units. The samples may well come from different basic data sources, and they will certainly emphasize different features. The basic household model may employ a uniform sampling ratio throughout; the taxpayer model may oversample the wealthy, and the benefit recipient model may oversample the poor. But they should produce the same distributions, and the distributions should agree with what the government model has to say about enrolments, costs and benefits, receipts and disbursements.

31. The government model may include a variety of kinds of information, so long as it refers to the governmental units upon which the model is based. The data may be financial: costs and benefits, receipts and disbursements, assets and liabilities. They may be physical: numbers of pupils and hospital patients, numbers of schools and hospitals, numbers of teachers and nurses, numbers of school lunches and surgical operations. They may deal with flows - annual expenditures or numbers of people served - or with stocks - value of hospital buildings and equipment or number of school buses. The essential principle is the possibility of attachment as an attribute of the individual reporting unit, in this case, the governmental organization.

32. It will be immediately obvious that there are many levels of governmental units, from the national Government down to the individual school or even classroom. The same problems of keeping straight the relationships among the different levels of the hierarchy enter here as in the household model, where they take the form of relationships among individuals, families, households and all of the other possible groupings of people.

33. In this connexion, mention may be made of the concept of the satellite account, a device which some countries have used. For Governments, satellite accounts have been used to accommodate additional information on aspects of government activity that do not fit conveniently into the aggregate government sector of the national accounts. In concept, satellite accounts are very similar to the nested submatrices of the demographic accounts and serve much the same function. A satellite account may, for instance, examine the educational system or the health-services delivery system. As usually employed, satellite accounts are thought of as further disaggregations of some part of the macro-economic accounting system. As such, they can also be accommodated in the over-all structure being considered here, as one of the levels of aggregation upon which it should be possible to enter the data structure. It should be possible to derive a satellite account relating to the health-care delivery system, for instance, by aggregation of information about the individual hospitals and other components of the system.

2. Establishments and enterprises

34. Social concerns, as noted above, are usually construed to include most aspects of economic activity, at least in so far as they relate to the activities of individuals, families and households. A large part of the population spends much of its waking hours in the environment of the workplace, and their economic activity constrains what they are able to do in their non-working time. Thus labour force participation, occupation, wages and other income shares, working hours and conditions, union contracts, fringe benefits - all are important for analyses whose main focus is social, as well as those dealing with more traditional economic topics.

35. Much of this information is, however, more readily collectable from the establishments and enterprises on the dispensing side of these relationships than from the workers and their families who are the recipients. Where the establishments and enterprises are large, or subject to government regulation, they are likely to have more systematic record-keeping procedures than are households. Such regular data-gathering exercises as the industrial census, and such administrative sources as tax authorities and factory inspectors, can provide information difficult to

obtain in other ways. For smaller enterprises and the traditional or informal kinds of production that are common in developing countries, these sources are less applicable and the household will remain the appropriate reporting unit; but even in developing countries, a large share of productive activity and a large part of the working population fall into the part of the economy about which establishment-oriented information is available.

36. Here again, however, there are multiple sources of data, and the familiar problems of achieving consistency among data sources and over time will arise. The compilation of an industrial register, a standard recommendation for the collection of industrial statistics, can play a vital co-ordinating role. Such a register identifies all of the establishments (or sometimes all those over a minimum size) operating within a country and specifies the relationships of establishments to their parent enterprises. It thus meets the need, already emphasized in connexion with both household and government data, for keeping straight the hierarchical relationships among related entities. It also can serve as the frame for a micro-analytic model based on establishments, by providing the means of connecting data from diverse sources. While it is often possible to treat small establishments in the same way one would treat households, relying upon samples and the construction of a model population of anonymous units, the larger establishments and their parent enterprises in most countries are too few, and too unique, to be handled readily in this fashion. Exact matching is usually necessary if meaningful results are to be obtained, and the industrial register is an extremely useful tool in this endeavour. Exact matching of identified units does, of course, create confidentiality problems. In most cases, there are some data relating to establishments and enterprises that are not considered to be confidential, and in some countries, efforts have been made to organize this publicly available information into micro-data base form. But much of the information about establishments, both that collected for statistical purposes and that arising from administrative sources, is unquestionably confidential, and developing ways to use it in disaggregated form without violating that confidentiality is essential.

37. As is the case with government data, establishment and enterprise data will also need to be related to household data. The principal instrument is, once again, the use of consistent definitions and classifications. The incomes paid out by enterprises should be defined and classified so that the same elements can be identified in the receipts of households. Employees of establishments, with appropriate occupations and qualifications, should be locatable among the inhabitants of households. Work-related injuries and disabilities should be traceable from the establishments where they occur to the households whose lives they alter. Changes in hours and conditions of work, the age and sex of the work force and the incidence of part-time work need to be related to the numbers and ages of children in the households concerned, the presence of other adults in the households, the availability of schools and their hours and alternative day-care facilities. For all of these kinds of analyses, it is essential to be able to move freely among data bases organized on the basis of different reference units.

3. Geographical or spatial units

38. All of the kinds of data discussed so far have dealt with actors on the social and economic scene: decision-making units of one sort or another. There is,

however, another important dimension that is needed for many kinds of analyses; it is is often important to know not only who and what but also where. Location is an attribute that can be, and often is, attached to the reporting units of household establishment and government data bases. But there are characteristics of localities and regions that pertain to the spatial units themselves and not just to the actors located within them, and it is often convenient to arrange such information on the basis of geographical or spatial reporting units. The kinds of information that might be included are physical descriptors, like area; features of the terrain, like water or mountains; climate, including rainfall, temperature, wind; political factors, like form of government; economic factors, like kinds of industry present; social factors, like availability of services. Some of this information clearly overlaps information arranged by decision-makers, and again preserving consistency is essential. Information on health-care facilities in a given locality in the geographically-organized data should agree with information in the government data. Information on the distribution of particular industries in the regional data should agree with the distribution emerging from the establishment data.

39. One special form of regional information that deserves mention is data about the environment. The environment is by its very nature spatially determined, and most of the topics usually considered to fall within the province of environment statistics require a regional dimension. Pollution originates from the activities of a given plant, mine, farm or city; it affects the quality of a given water- or air-shed. Land use and human settlements are not general concepts but relate to particular places. Environmental services are used by consumers and producers in a regional context, and environmental damages are place-specific. There are some environmental considerations that are national or even world-wide in scope, but these are rare. Improving the air quality in locality A is usually no help to locality B, and there are few general measures that affect both unless they are geographically close. The environment does, nevertheless, play an important role in both economic and social questions, and linkages are therefore necessary between the environment data and the data about households, Governments and enterprises. It is, for many kinds of analysis, important to be able to identify who uses environmental services, who suffers from environmental damage and who pays the costs involved. Manufacturing plants that use river water for cooling or for waste disposal; cities that discharge sewage into the nearest body of water; farmers who overgraze range lands or strip the land of trees for fuel; automobile drivers and electricity-generating plants that burn fuels high in pollutants are all obtaining services from the environment. Usually, the users of the services do not pay for them, either in the sense of compensating those who suffer from the damage caused or in the sense of making the damage good. Increasingly, however, it is being recognized that environmental assets like clean air and water are not really free and that natural resources, traditionally treated like a gift of nature, are really quite exhaustible. Data relating to the environment are a necessary part of the integrated demographic, social and economic data structure.

D. Convergence of the classificatory framework for economic and social fields

40. It is not coincidence that the kinds of organizing principles and reporting or reference units identified in paragraphs 29-39 above as relevant for social

analysis are the same as those identified as sectors in the macro-economic national accounts. Social analysis is concerned with people and their behaviour and interaction, but the setting in which they find themselves is an important determinant of how they function, and it has a critical bearing upon the development of policy and the evaluation of results. It is not possible to understand the forces at work or the mechanisms through which they operate without taking all of these considerations into account. It is thus becoming clear that a unified data structure is needed that can accommodate the data requirements of both economic and social analyses.

41. The national accounts have an important role to play in such a structure as an organizing device, not only for economic statistics but for social statistics as well. "Organizing device" here does not mean straitjacket, nor does it imply the subordination of social to economic concerns. What it does mean is that the same sectoral framework is appropriate, that many of the same classifications and definitions are useful and that one way the social and demographic micro-data must add up is to the macro-economic accounts. There are other ways, and other aggregate structures, and the ultimate over-all data structure must encompass such alternatives, but it must do so in such a way that the interrelationships are clear. A demographic account concerned with the flows of the population through the educational process, for instance, should bear a clear relation to the household micro-data model, on the one hand, and the government micro-data model, on the other, and these in turn should bear a clear relation to the household and government sectors of the national accounts.

42. In order to make such interrelationships possible, adjustments and accommodations may be needed in the national accounts. Some initial steps towards the needed kinds of accommodations were recognized in the work on the distribution of income undertaken in connexion with the development of SNA. The publication entitled Provisional Guidelines on Statistics of the Distribution of Income, Consumption and Accumulation of Households 7/ includes a table showing the adjustments required to move from the concepts of SNA to aggregates that do represent summations of household data, and several new concepts are introduced that better reflect the processes of income creation, distribution and disposition as they occur in individual decision-making units. 8/ The kinds of questions that arise are explored further in the report on SNA mentioned in paragraph 28 above. For the most part, the adjustments are not only needed for integration of social and economic data, but would improve the usefulness of the accounts for purely economic analysis.

E.   Complementarity of macro-data and micro-data structures

43. The unified information system that is envisaged in the present report will need to accommodate not only data on both economic and social topics, but data on all levels of aggregation. Macro-data and micro-data should not be thought of as

_____

7/ United Nations publication, Sales No. E.77.XVII.11 and Corr.1.

8/ Ibid., annex I, table 1, and chap. II, table II.1.

alternatives, but as complements. Each provides necessary support and supplementation of the other; together they are a much more powerful tool than either separately.

44. Micro-unit-level data provide support for the aggregate level data in a number of ways. First, a micro-data base is an efficient way to store the data required to obtain more detailed disaggregations of the macro-data. The distribution of income, for example, can more easily be obtained by aggregation of an appropriate micro-data base that is aligned with the relevant aggregates than by more traditional methods of disaggregation.

45. Secondly, micro-data provide a simple means of separating the changes observed in the macro-data into structural and behavioural components. A change in the aggregate level of income, for example, may arise either from widespread changes in the incomes of all or most households or from a change in the distribution of households between high-income and low-income groups. A special but important case of the latter is rural to urban migration. Where an observed increase in aggregate income comes about solely because of the increased income of urban in-migrants, with little change in the incomes either of those remaining in rural areas or those who were already in the cities, the policy response called for might be quite different from that which would be appropriate if the increase in income affected all groups in the society more or less equally.

46. Thirdly, underlaying the macro-data with appropriate micro-data models provides a convenient way to link the data observed at the macro level to characteristics or attributes that are not easily aggregated, or for which the numeraires employed in the aggregate data (monetary units, people etc.) may be inappropriate. To continue the income example, one may wish to know the relation of household income to family composition, location or socio-economic categories. These are easily handled as attributes of individual households. This aspect is particularly important when consideration is given to the directions towards which future policy and research interests are tending. On the one hand, study of the household as a producing and consuming unit is being pushed back, from its function as a producer of marketed labour services and a consumer of GDP-type goods and services, to a study of the whole life environment, including the use of time by all of the household members, their own evaluation of the "process benefits" derived from the activities they engage in and subjective measures of the satisfactions they derive from these activities. On the other hand, the range of factors whose influence upon the household is being considered is being broadened to include, for instance, the physical environment.

47. Conversely, the macro-level framework is equally necessary as a support for micro-level data. First, it provides the framework within which the micro data fit and makes it possible to obtain an overview of the whole picture, thus placing a particular micro analysis into appropriate perspective. It is helpful for policy-makers dealing with a particular client group, for instance, to know how important that group is in the population as a whole and in various segments of it. What proportion of the total number of households is below the poverty level? What proportion of the households below the poverty level have no gainfully employed member, and what proportion of the gainfully employed are in poverty? The approach

to poverty will be quite different if poverty households are 80 per cent or 20 per cent of the total, if they are mainly unemployable, mainly economically active but unemployed or mainly employed but at jobs that do not yield a living wage.

48. Secondly, macro data can provide convenient summaries of important aspects of the micro data at intermediate levels of aggregation. The satellite accounts of the national economic accounts and the nested submatrices of the demographic accounts, as already noted, serve this function. Such an intermediate level aggregation may be used, for instance, to examine in more detail than the aggregate systems commonly have room for such questions as the financing of the health-care system or the incidence of violent crimes and the characteristics of the victims. The subsectoring of the household sector introduced into the World Bank's Social Accounting Matrices (SAMs) also falls into this class; subgroups of the population such as farm households, the aged or households below the poverty level can be examined in this way.

49. Thirdly, the macro data provide constructs that may themselves be important elements in the micro analysis. The consumer price index, for example, or aggregate measures of the change in productivity or the change in the birth rate may have an impact upon the circumstances and behaviour of individual micro units, either through such legally mandated links as escalation clauses in wage contracts or indexation of social security benefits, or through policy responses to macro statistical triggers. In this connexion, macro-level social indicators are an important statistical output, and in the future they may be expected to include macro-level counterparts of the new concerns being reflected on the micro level, such as time use and subjective measures of well-being.

50. From the point of view of the development of the integrated information framework, it is necessary to specify and design both the macro systems and the micro systems in such a way that they form an integrated and articulated whole. Unfortunately, many of those concerned with the development of macro systems (both economic and demographic) have failed to consider the nature of the micro-data systems which underlie them and upon which they must conceptually rely. Conversely, those concerned both with household surveys and with the collection of data from firms and governmental units have often not considered the macro-data systems into which the micro data must fit. The importance of developing a fusion of macro analysis and micro analysis is increasingly being recognized; for this to be possible, a prerequisite is integration of the data.

51. The required integration of macro-level and micro-level data is in practice gradually coming about, less by design than by necessity. The greatest progress is being made where the data are needed for specific policy purposes, mostly concerned with measuring the benefits and costs of specific government programmes. Statistical offices have not been the moving force in these developments, which have often been undertaken by policy users rather than producers of the statistics. The time has come when the further development of the structure should be considered explicitly by the producers of statistics in order to ensure that its growth is rational and directed and not left to chance and the accidents of funding.

/...

III. FUTURE WORK

A. Work programme adopted in 1979

52. At its twentieth session, the Statistical Commission considered a programme of future work in this area for the Statistical Office, which had been drawn up by an expert group convened in 1978. 9/ The programme was an extensive one, and it was recognized by both the expert group and the Commission that it was far beyond what it would be feasible to carry out with the resources available to the Statistical Office. In the event, resources have proved to be even more limited than expected, owing to staff turnover and recruiting difficulties. Nevertheless, it may be useful to outline briefly the content of that programme and to review the conclusions of the Commission as to priorities before discussing the progress made and future work.

53. The programme was divided into four parts. The first dealt with subject-matter questions and included work on concepts and classifications for specific social fields, a study of the linkages among different fields and an examination of analytic models employing social and demographic data. The second part dealt with organizational and procedural problems of data collection and management, including questions of documentation and archiving, confidentiality and methods of obtaining access to administrative data. The third part dealt with questions of statistical methodology that are particularly involved in the construction of micro-data bases and their integration with aggregated data. The topics included were techniques of estimation, the logical design of micro-data bases so as to permit maximum flexibility of access and use, and problems of effective use of administrative data. The final part of the programme dealt with computer technology, in so far as it is peculiarly related to the production, storage and dissemination of social statistics focusing in particular on the preparation of manuals and technical reports.

54. The Commission considered that priority should be given to the first and third parts of the programme and, in particular, to: (a) the development of concepts and classifications; (b) consideration of the structure of linkages; and (c) statistical methodology, especially techniques of merging micro-data bases. During the last two years, the Statistical Office has devoted its efforts primarily to (a) with some attention to (b). Resources have not been available to proceed to (c).

55. The present section of the report will review the work done in the last two years, setting it into the context of a discussion of what is immediately needed in order to make it possible to move towards the objective of the establishment of a general information system. As was noted above, this goal is one that can be approached on a gradual, step-by-step basis. It relies, of course, primarily upon the efforts of national statistical offices; the contribution of the United Nations Statistical Office is of necessity limited to the suggestion of ways and means. For the Office, priorities remain the same as those identified by the

---

9/ Official Records of the Economic and Social Council, 1979, Supplement No. 3 (E/1979/23), paras. 115-117.

Commission in 1979; the section below attempts to indicate how the work done by the Office may be utilized by countries in taking initial steps towards the establishment of a general information system.

B. Starting point and immediate uses

56. In practical terms, even though the ultimate objective is a system of organizing micro data and macro data that will provide a place for whatever social and economic data a country considers to be useful, the construction of such a data structure must make a start in some particular area. From the point of view of social and demographic statistics, the place to start, as already suggested, is with the construction of a micro-data base that is a model of the population. This requires a definition of the structure of the model, including its relation to suitable aggregate demographic accounts and to the household sector of the national economic accounts and the provision of appropriate definitions and classifications.

57. The development of such a household micro-data model integrated with the aggregate demographic and economic accounts would yield immediate direct benefits. It would make possible the subsectoring of the household sector according to socio-economic categories (farm, nonfarm entrepreneurial, nonfarm wage earner, other) as is recommended in SNA. It would make possible the construction of income distributions along the lines recommended in the Provisional Guidelines on Statistics of the Distribution of Income, Consumption and Accumulation of Households 10/ and their extension to accommodate aspects of levels of living not easily reducible to monetary measures. It would make possible the construction of a variety of social indicators relating to health, education, the quality of life and levels of living and the impact of environmental factors. It would permit the analysis of the impact of government tax measures and benefit programmes, individually and in the aggregate, upon households of different characteristics. It would provide a framework for analysing the learning and earning sequence.

58. The National Household Survey Capability Programme can be expected to make a major contribution to the availability of suitable micro-data for the construction of such population models but, in order to ensure that the survey data collected are useful for this purpose, a specific focus is needed on the creation of reusable micro-data sets aligned with both the demographic and economic macro data.

C. Concept of the specialized subfile

59. For any particular application, it is quite likely that one would not want to make use of the entire model population, any more than one would expect to use all of the national accounting data in a particular macro-analytic application. In such a case, a specialized subfile containing only what is useful for a particular

10/ United Nations publication, Sales No. E.77.XVII.11 and Corr.1.

problem can be created. This can be done in a variety of ways. It may be possible
to extract the required information from the general model. Or, especially in the
early stages of development of the general model, it may be more efficient to create
a specialized file through a special study. But whatever the source of data
and method of subfile construction used, it should be possible - as it is for the
macro-data - to relate the specialized information that is needed for a particular
problem to the general model in a way that preserves the consistency of the
specialized data with the general information system. This would mean that, for
the attributes which both the general model and the specialized subfile have in
common, the aggregates and distributions that could be generated for any similarly
defined group of individual units would be directly comparable. Thus, for
example, if a special subfile were created to analyse children's health, the
sample should conform in age, sex, ethnicity and locational characteristics and
any other common attributes with the cases observed for children in the general
model. This does not imply that the specialized subfile must be obtained by
extracting data from the general file or that the same sampling ratios must be
used. As already suggested, many problems may require the use of stratified
sampling for the data to be analytically useful.

60. Depending upon the use in view, such a subfile may be specialized by region,
topic, client group etc. Thus, it may not include representatives of all of the
types of units present in the general file. In addition, it may not be necessary
to include in the specialized subfile all of the attributes of the micro units
that are available in the general file. For use in studying child health, for
instance, information on crime victimization or voting participation would
probably not be of great utility. On the other hand, it might well be desirable to
include attributes relating to the particular problem at hand (for instance,
detailed data on morbidity and mortality) that are too specialized to warrant
inclusion in the general model. There is, of course, no barrier to the inclusion
of additional attributes in specialized subfiles, so long as the additions are
made in ways that preserve the known relationships between the attributes being
added and those already included and so long as alignment with known control
totals is maintained.

D. Reporting and reference units

1. Possibility of hierarchical classifications

61. Various kinds of reporting and reference units may be encountered in the
collection and analysis of household data. Different sources of data employ
different reporting units, and different kinds of analyses will require different
reference units. Households may contain several families, or families may be
spread over several households. Families may contain subfamilies, or they may
be part of extended families. Individuals will enter into all of these, and there
may in addition be various other kinds of units such as tax reporting units and
units receiving the benefits associated with particular government programmes.
The choices made among the various possible units, their definition and
classification and the specification of the nature of the relationships among them
are of primary importance for the construction of a population model that is
intended as a general-purpose analytical tool.

/...

62. It is not fruitful to look for one "best" unit or for a definition that will serve all purposes. What is needed is, first, to identify the kinds of units that are commonly encountered and, secondly, to determine which of these units can be defined and classified in ways that will clearly specify the hierarchical relationships among them. What units should be included and what the relationships are may well differ from country to country, and there are limits to the degree of international comparability that it is useful to strive for. First efforts should focus on specification of the units and their relationships within a given legal and cultural context; it may then be possible to consider what degree of commonality can be found on a regional or world-wide level. The Statistical Office has in the past specified only one concept – the census household – for international reporting. The Office is now engaged in exploratory study of the possibilities for development of a more extended hierarchical structure of household concepts, together with appropriate classifications of them, both for international reporting purposes and for use by countries in meeting their own national needs.

2. Methods of integrating non-hierarchical units

63. There will, however, always be special-purpose reporting units and groupings of individuals that will not fit into such a hierarchical structure. Tax reporting units, for example, may or may not coincide with any of the units included in the hierarchical structure. Units referring to the recipients of the benefits of particular governmental programmes almost certainly will have their own special definitions. Relating data based upon such special units to data using more general household units has always been one of the more difficult aspects of integration of data from different sources. The problem is very considerably simplified when the data are stored in micro-unit form. For any given household, the link between such special groupings and the groupings of the hierarchical structure can be made through the individuals that are contained in each. Data based upon one reference unit – for instance, the census household – can be converted to, say, income-tax units at the level of the individual micro-unit, so that in turn tabulations on one basis can be converted to another basis by retabulating the converted micro-data. Unlike the problems of developing hierarchical structures and classifications, problems of this sort do not lend themselves to international standardization, since of course, they depend upon particular legal and institutional arrangements. What can usefully be examined from an international point of view, however, is the methodology involved. The conclusions in any particular case will depend upon the circumstances of the country, but the procedures required are likely to be very similar in different countries.

E. Attributes

1. Nature of classification systems

64. The use of common definitions and classifications of attributes of the micro units is what holds the general model and associated specialized subfiles together and makes it possible to move freely among the specialized files. The definitions and classifications of the attributes are, therefore, of critical

importance. This point is, of course, one whose importance has been emphasized in earlier reports, and there is no need to repeat the earlier discussion here. It may, however, be useful to outline the various classes of attributes to which attention must be devoted and to indicate the present status of work on them. In most of the areas indicated below, international work has been initiated; in many of them, standards have been adopted. Each individual country needs to consider these international standard concepts and classifications from the point of view of their suitability for the country's own circumstances. While adherence to the international standards is to be strongly urged for international reporting purposes, it is to be expected that countries will wish to make modifications and adaptations for their own internal use. The main consideration is that decisions should be taken, and the definitions and classifications decided upon should then be applied as widely as possible to all statistics.

65. Even this, however, is a goal that can be approached gradually. One of the advantages of starting the process of integrating demographic, social and economic statistics with the construction of a model population is that it is not necessary to obtain universal compliance immediately. Where data collected and used for special purposes are based on non-conforming definitions and classifications, they can be left undisturbed until the usefulness of consistency has been demonstrated. Even in the longer term, it is likely that there will always be some special needs that can only with difficulty be fitted into the standard definitions and classifications and, as new topics arise and assume importance, time will be required to develop standards for them. The objective is to extend the coverage of the generalized information system and related specialized subfiles as broadly as possible, as the opportunity and possibility of doing so arise.

66. The classification system should, furthermore, not be considered to be a static thing, but as something that grows and changes to meet changing circumstances and needs. The need for periodic review and updating of classifications has always been recognized in the case of the older international standard classifications such as the Standard International Trade Classification (SITC) and the International Standard Industrial Classification of All Economic Activities (ISIC); as technology changes, new industries arise and old ones disappear, and the classifications must reflect the actual situation. As experience accumulates with the newer classifications, it is to be expected that they, too, will need updating and modification to reflect the changing face of society.

2. General attributes

67. There are a few classes of attributes that are likely to be wanted in any household micro-data set concerned with demographic, social and economic data. For each individual included in the reporting unit, one wants at least age and sex. For the reporting units as a whole, one wants information on the individuals included, the various subunits of which each is a member and their roles in the subunits. Finally, one wants spatial identification: where is the household located and where do its members work?

68. Reports on classifications pertaining to two of these topics, age and location,

/...

are before the Commission under agenda items 7 (e) and (f). With respect to age, "Draft guidelines on age-group classifications" (E/CN.3/550) gives information on national practices relating to age classification in data collection, storage and presentation, current international age classifications and recommendations for international classifications for 12 subject areas at three levels of detail. The classifications recommended have been designed not only to meet the needs of the specific subject-matter areas but to facilitate the comparison of information across subject-matter fields, and their use would greatly improve possibilities of bringing aggregated data from different fields together. As the report points out, for the basic micro data it is to be preferred that single years of age be collected and stored. With respect to spatial identification, work has not progressed as far. The report, "National practices in classifications of size and type of locality and urban/rural areas" (E/CN.3/551), contains a discussion of the conceptual issues involved and of national practices on such questions as the definition of urban and rural areas, the identifiability of localities by size of population, the determination of locality boundaries, population density and types of geographical units for statistical purposes. Proposals are made in the report for further work directed towards the development of guidelines for standard classifications of geographical units combined with a classification of urban and rural areas.

69. The third general area relates to the definition and classification of households and related units. As noted in paragraph 62 above, the Statistical Office has begun an exploratory study on this topic. As is widely recognized, this is an area of considerable conceptual complexity and wide diversity among countries. If the Commission wishes, a progress report on the work can be prepared for its twenty-second session.

3. Framework for social and demographic statistics: concepts and classifications

70. The status of work on classifications relating to the various special fields of FSDS was summarized in Part Three and annexes I, II and III of Studies in the Integration of Social Statistics: Technical Report. 11/ As is noted there, the development of international standard classifications has been of major concern to those working in the special fields, and in most areas, international recommendations have been adopted. It is entirely appropriate that the major task of developing and maintaining classifications in the special fields should be the responsibility of the specialized agencies working in each of the fields. From the point of view of the construction of a generalized information system, however, it is essential that all of the specialized classifications be examined in terms of their usefulness not only within special fields but as the means of establishing bridges among those fields. A classification of diseases and causes of death, for instance, may be needed by those looking into such diverse topics as industrial working conditions and injuries, highway accidents, crime victimization, school attendance and performance and household budgets, as well as by those concerned with questions more narrowly relating to health. The classifications must be constructed in a hierarchical fashion that permits their use, at perhaps a less detailed level, outside of the area of their principal focus; only in this way will they serve the function of knitting together the whole fabric of the general information system.

11/ United Nations publication, Sales No. E.79.XVII.4.

/...

4. System of National Accounts: concepts and classifications

71. Attention also needs to be devoted, from a general information system point of view, to economic concepts and classifications. Some of the economic classifications are of direct and immediate concern in constructing a population model, and some are more peripherally related, but all need to be considered. The economic classifications have, in general, a longer history, and many of them have been brought together under the aegis of SNA. Nevertheless, there is active work going on in a number of fields. The work of the Statistical Office is reported on under agenda item 6, in "Progress report on harmonizing economic classifications" (E/CN.3/545). The main activity reported there is the work on the revision and updating of ISIC being undertaken with the advice of the Joint United Nations Statistical Office/Statistical Office of the European Communities Working Party on World-Level Classifications. The discussion need not be repeated here, save to point out that one of the objectives of the revision is to improve its usefulness as a co-ordinating device in relating production, trade and consumption data.

72. A new Classification of the Functions of Government has been completed. 12/ This replaces the classification appearing as table 5.3 in SNA. The new version provides for updating and further detailing, and also for some adjustments and modifications that experience suggests would improve its usefulness for such purposes as tracing the benefits of various government programmes and compiling such aggregates as total consumption of the population. In preparing the new classification, efforts were made to co-ordinate it with related work of the International Monetary Fund (IMF) and the Organisation for Economic Co-operation and Development (OECD). In the same general area of public sector statistics, IMF and OECD have been working on a classification of the receipts of Government.

73. Active work is also under way, primarily in the International Labour Office (ILO), on concepts and classifications relating to labour, including employment, unemployment and underemployment; the concept of economic activity; and the classification of types of activity status. It has become clear, in the interval since the current definitions and classifications were adopted, that the questions at issue are more complex than they were earlier seen to be, and provision needs to be made both for situations and circumstances not then envisaged and for a closer relation to such topics as time use and non-market production activity. The Statistical Office expects to contribute to this work, especially in terms of repercussions outside of its immediate area of concern.

74. A final area that should be mentioned in this context is that of the classification of private consumption expenditures. The System of National Accounts includes a classification (table 6.1) of household consumption expenditures intended for use in compiling the national accounts, and countries have widely adopted the classification. It is not, however, nearly as widely used for such related purposes as compilation of the consumer price index, with the consequence that it is difficult to relate expenditure and price data, or often to relate the data in the macro-national accounts to data obtained in household income and expenditure

---

12/ Statistical Papers, Series M, No. 70 (United Nations publication), in press.

surveys. These questions need further examination with a view to updating the classification and improving its applicability for wider use; in this connexion, its compatibility with the International Classification of All Goods and Services (ICGS) needs to be explored.

5. New areas

75. As new kinds of statistics are developed, an essential part of the process is the formulation of concepts and the construction of classifications. Work in the Statistical Office is at this stage in one important new area, namely statistics of the environment. A "Progress report on environment statistics" (E/CN.3/553) is before the Commission under agenda item 8. Among other topics, it discusses the preparation of technical reports on the environmental concerns affecting human settlements, land, energy, natural resources and pollution; and the development of a preliminary format and structure for the organization of data in this field. Work on concepts and classifications relating to environmental topics is of growing importance for ensuring that, as environment data come to be more widely available, they are compatible with the social and economic data with which they must be used.

76. There are also various other newly developing areas of social statistics where work on concepts and classifications is needed, including statistics of time use and subjective evaluations of well-being. The work of the Statistical Office in these areas has been limited to a general monitoring of country practices and of the work of other international organizations such as OECD. A report on time-use statistics was presented to the Statistical Commission at its twentieth session (E/CN.3/519), and Social Indicators: Preliminary Guidelines and Illustrative Series 13/ includes a discussion of the OECD work on subjective indicators. Further work in both of these areas is urgently needed.

F. Statistical methodology

77. The final priority area for further work identified by the Statistical Commission (see paras. 52-54 above) was statistical methodology relating to the construction of micro-data bases and their integration with aggregated data. This was construed to include (a) techniques of editing and imputation of missing values, exact and statistical matching, problems of alignment and reconciliation of data from different sources and problems of statistical quality control; (b) the logical, conceptual aspects of constructing micro-data bases so as to permit maximum flexibility of access and use at various levels of aggregation; and (c) problems arising in making effective use of administrative data, such as conversion from one type of unit to another, consistency of coverage, etc.

78. It is apparent that all of these questions retain their urgency. Effective solutions are vital to the conception of a general information framework outlined in the present report. The Statistical Office has been unable to date to initiate work in these areas owing to lack of resources. For most of these questions,

---

13/ United Nations publication, Sales No. E.78.XVII.8.

specialized expertise is needed. The most effective procedure is, therefore, likely to be the preparation of a series of technical reports by consultants engaged for this purpose. There are a number of topics on which such reports could be extremely useful in consolidating national experiences and setting out recommended procedures. They might include:

(a) Editing, imputation and alignment;

(b) Statistical matching of micro-data files: methodology and tests of validity;

(c) Uses of administrative data;

(d) The design of a model population; and.

(e) Linking of specialized micro-data files.

G. Summary of proposed work programme

79. The proposed work programme may be summarized as follows:

1. Structure of linkages among fields

Continuation of the work represented by the present report and the companion report on SNA (E/CN.3/541) on the relationships between macro-data and micro-data systems and the design of suitable macro systems, including demographic accounts. Consideration, as a part of the work on SNA, of the implications of the conception outlined in the present report for the future development of the national accounts.

2. Classifications and definitions

(a) Continuation of work in the areas already identified as Statistical Office responsibilities, including geographical classifications, household concepts and classifications and environmental concepts and classifications;

(b) Co-operation with groups working on specialized topics with a view to ensuring cross-field compatibility;

(c) Initiation of work in new fields such as time use.

3. Statistical methodology

Preparation, by consultants, of technical reports in the five areas outlined in paragraph 78 above: editing, statistical matching, use of administrative data, the model population and linking of specialized subfiles.

IV. POINTS FOR DISCUSSION

80. The Commission may wish to:

(a) Express its views on the approach to the integration of demographic, social and economic statistics presented;

(b) Consider whether a technical report on this topic should be published, incorporating the comments of the Commission; and

(c) Comment on the work programme proposed, especially with regard to priorities (para. 79).