



Economic and Social Council

Distr.
GENERAL

E/CN.3/1983/3
23 September 1982

ORIGINAL: ENGLISH

STATISTICAL COMMISSION
Twenty-second session
7-16 March 1983
Item 3 (b) of the provisional agenda*

SPECIAL ISSUES: INFORMATION TECHNOLOGY

Implications of new developments in information technology

Report of the Secretary-General

SUMMARY

The present report begins with an outline of information technologies relevant to the work of the statistician in the fields of computer hardware and software (paras. 9-22). A summary of current levels of usage and trends is then presented (paras. 23-53). An indication of future possibilities and impact is given (paras. 54-103). Points for discussion by the Commission are included (para. 104).

* E/CN.3/1983/1.

CONTENTS

	<u>Paragraphs</u>	<u>Page</u>
INTRODUCTION	1 - 8	4
I. INFORMATION TECHNOLOGIES RELEVANT TO THE WORK OF THE STATISTICIAN	9 - 22	5
A. Hardware	10 - 17	5
B. Software	18 - 22	6
II. PRESENT LEVEL OF UTILIZATION OF VARIOUS INFORMATION TECHNOLOGIES	23 - 53	7
A. General experience and trends	23 - 30	7
B. Trends and experience in official statistical offices of developed countries	31 - 38	9
C. Trends and experience in official statistical offices in developing countries	39 - 44	10
D. Trends and experience in international statistical offices	45 - 49	11
E. Trends and experience outside of official statistical systems	50 - 53	12
III. FUTURE POSSIBILITIES FOR EXPLOITING INFORMATION TECHNOLOGY IN A STATISTICAL ENVIRONMENT	54 - 101	13
A. Statistical networks	54 - 69	13
1. National networks	55 - 58	13
2. International networks	59 - 69	14
B. Impact on economic and social development	70 - 73	15
C. Impact on the production of statistics	74 - 93	16
D. Impact on the users of statistics	94 - 96	19
E. Impact on respondents	97 - 101	20
IV. SUMMARY	102 - 103	20
V. POINTS FOR DISCUSSION	104	21

CONTENTS (continued)

Annexes

- I. POSSIBLE WORLD STATISTICAL NETWORK
- II. EXISTING UNITED NATIONS COMMUNICATION SYSTEM

INTRODUCTION

1. It is now little more than 30 years since the first large-scale electronic computer available commercially, the UNIVAC I, was installed. The very first model was installed in the United States Bureau of the Census. It was a decimal, alphanumeric machine with 12,000 characters of high-speed primary storage (200 microseconds to access a 12-character word). A few other countries began installing computers in the 1950s, the first one very often in the central statistical office.
2. In 1962, the Statistical Commission first considered documents on data processing in response to a decision at its eleventh session to consider as a continuing project of high priority the study of the methodology of automatic data processing. 1/
3. In the 20 years since then, almost all national and international statistical offices have been computerized. The computers being used have grown tremendously in capacity and power per unit cost. For example, compared to the UNIVAC I, occupying hundreds of square feet and costing over \$2 million, one can now purchase a desk-top computer for \$2,000, with more memory, a hundred times faster access and five hundred times more powerful. This kind of explosive growth has had an enormous impact on statistical services.
4. At its twenty-first session, the Statistical Commission decided to consider the impact of technology as part of the more "future-oriented" content of its twenty-second session. 2/ It was considered that both the demand for statistical information and the technological capabilities for processing it were changing so rapidly that the environment for statistics, at both the national and international levels, would be affected.
5. The present report is a broad summary, on the one hand, of current practices and trends and, on the other, of a set of postulates with respect to future developments and their potential implications for all aspects of official statistics.
6. Chapter I provides a brief list of information technologies considered to be appropriate to the processing of official statistics. It is not intended to be an exhaustive or technical list, but rather an indication of the more important features of the technologies employed - both the traditional and the newer ones. In chapter II, an attempt is made to summarize the ways in which these technologies are being used. Because the situation is somewhat different in developing and developed countries, they are treated separately to a certain extent. Some general observations and experiences are noted in the conclusion to the chapter, which is intended to summarize the current situations and trends on which chapter III is based.
7. In chapter III, future possibilities and implications are examined. The chapter is inevitably somewhat speculative, but it is felt that, if adjustments in the statistical systems are going to be demanded, then discussion and planning - and even speculation - must take place now.

8. In preparing the present document, note has been taken of the many reports devoted to this subject in the last decade at the international level. These included the report prepared for the Conference of European Statisticians; 3/ the report of the ESCAP Study Group on the Co-ordination of Government Information Systems; 4/ the proposals for Data for Development to conduct studies in this area; 5/ the report of the Seminar conducted by the Conference of European Statisticians in Moscow in September 1981; 6/ and various papers presented at the forty-third session of the International Statistical Institute at Buenos Aires in November 1981. 7/ 8/

I. INFORMATION TECHNOLOGIES RELEVANT TO THE
WORK OF THE STATISTICIAN

9. A list of the various information technologies appropriate to statistics is given hereunder in as non-technical a fashion as possible.

A. Hardware

10. For the purposes of capturing source data, the punched card is rapidly disappearing. Key-to-disk or key-to-diskette has become the standard technique. One special development in the last decade has been the grouping of data-entry terminals around a small computer. In this way, the checking, correction and imputation activities can be programmed for these small computers which, in turn, can be distributed geographically.

11. Other forms of source data automation are very slowly being introduced. Mark-sense readers are being used increasingly in large surveys and census work. However, there has been little growth over the last decade or so in the use in statistical offices of OCR (Optical Character Recognition).

12. In the architecture of the central processors themselves, the last decade has seen the explosive growth of LSI (Large-Scale Integration) technology. This has made possible three very important advances: greater reliability, great reduction in size and sharp decreases in cost. One of the very important consequences has been the growth in the use of microcomputers and minicomputers. The power and flexibility of these devices continue to improve greatly, whilst the cost decreases.

13. The cost of mass secondary storage has decreased greatly. Whereas with the UNIVAC I only 12,500 words could be held in secondary direct-access storage, millions of records can be held in typical direct-access devices today. This capacity not only affects the amount of data that can be processed, but affects the design of techniques for processing them.

14. In the field of communications, there is tremendous activity. Large companies are installing their own private networks. The public networks of telephone and television facilities are being increasingly used to collect and disseminate information. As they are connected to the array of satellites now orbiting the earth, these offer vastly greater scope for international communication than was

possible a decade ago. In fact, the capacity of the satellite communications network has grown to such an extent that today one can transmit millions of records to anywhere in the world in minutes. A number of so-called information utilities, disseminating different types of technical information, are now offering a world-wide service.

15. As regards other hard-copy output devices, the statistician still relies heavily on magnetic tape and the high-speed printer, technologies dating back to the 1950s. However, more recent technologies are becoming very important. First, the digital plotters have been refined greatly in the last five years. They are now much faster, more reliable and capable of producing graphic output in several colours. This refinement is slowly being reflected in national and international statistical publications.

16. Another instrument for producing graphic output, and one which is more productive and flexible than the plotter, is the graphic display terminal. These terminals can also produce results in several colours and, coupled with sophisticated software for graph formation, mapping etc., can produce high-quality output with much greater impact than conventional tables. A particularly impressive example is Social Indicators III of the United States. 9/

17. For the publication of statistics, statisticians have long relied on offset reproductions of computer printouts. These have improved greatly in quality. But for "publication-quality" material, the statistician has generally had to rely on external typesetting. Computerization of the typesetting process began 20 years ago, but only a massive volume of work justified the installation of in-house facilities for this technique. More recently, plates for printing have been prepared by computerized photocomposition. This technique has been developed rapidly in the last decade and, as with much of computing technology, it has become much cheaper. Some national and international statistical offices have now installed in-house photocomposition devices, driven directly by magnetic tapes from the computer.

B. Software

18. Significant improvements continue to be made in the design and availability of generalized programs for statistical data processing. A large number of table generators, operating on a variety of machines, are now available. A smaller but growing number of packages is being developed for data editing and error correction.

19. There are also some very powerful, easy-to-use and comprehensive statistical analysis packages available from commercial sources, as well as a sharing of programs developed in academic and government institutions. One very important development is improved portability, the systems being written in compiler languages available on a variety of machines. Some very useful software is even available on very small computers, even minicomputers.

20. One field of statistical software which is now beginning to receive some attention is that of computer-assisted survey design and maintenance. The

development of various registers (covering people, land and economic enterprises) opens up a much greater possibility of sample surveys. The proliferation of survey design has led to the search for computer assistance and this is beginning to materialize. In fact, the whole process of survey preparation, from design through forms preparation and dispatch to response control - all steps previously performed outside the computer - is now susceptible to computerization through the use of appropriate software.

21. The same is true of the other end of the processing spectrum - output. In traditional statistical computing systems, the printed table was the end product. Today, the design of the system relegates printed tables to a secondary position, or even abolishes them. Instead, thanks to advances in software and hardware, it is possible to photocompose the results for printing, producing graphs, preparing microfiches and storing the data on disk for on-line access by users using terminals.

22. In the general field of statistically oriented software, however, it is in the middle ground that the most important evolution has occurred, namely, in the development and use of data-base management systems. There have been three major ways in which this has taken place:

(a) the in-house development of a data-base management system (DBMS) especially suited to the data types, transactions, query processing, output needs etc. of the particular statistical agency;

(b) the adoption by one agency of such a system developed by another;

(c) the installation of a generalized data-base management system (GDBMS) acquired from another source, usually commercial.

Some agencies have already tried more than one approach. An evaluation of their experience is given in chapter II.

II. PRESENT LEVEL OF UTILIZATION OF VARIOUS INFORMATION TECHNOLOGIES

A. General experience and trends

23. The background of the present chapter is that of the growth of the use of computers in the last decade or two. The computer industry is the fastest-growing major industry in the world. It has been estimated that the value of sales of computers has risen from \$339 million in 1955 to over \$60 billion per year in the early 1980s. ^{10/} The computer has expanded the capacity of statistical offices to collect, manage and disseminate statistics.

24. In the early years after the introduction of computers into the official statistical process, much thought went into the definition and refinement of the processes. As a consequence, the processes were better understood and the quality of the results improved. The discipline of computerization in many areas has

changed the nature of the work, standardized it, made it of more even quality. The power of computing has led to a greater amount of checking, more sophisticated estimation, more extensive manipulation, derivation, calculation and tabulation and more flexible, selective and extensive dissemination of data.

25. As the need for more efficient methods of processing data grew, and once the computer was applied to structure and organize data, it was used to manipulate and explore the data in ways that created new information. Thus, although statisticians originally used the computer in order to organize information, it has grown into a machine which can be used as a tool for converting and providing information itself. A good example is the use of administrative records for statistical purposes. 11/

26. Modern societies have become increasingly dependent on computers and on information. Not only are more data being generated, but much greater use is being made of the information. Statistics are being gathered concerning new fields, such as the environment, social aspects, energy etc. The structure of societies - in both the public and private sectors - has become more complex. The range of goods manufactured has become greater and more complex. Governmental and private organizations are becoming more diversified. Conditions are changing rapidly and dependence on forecasting and modelling is growing.

27. All these factors have increased the demand for official statistics, and it is the computerized information system which has made possible a proper response. There are simply no other means available for satisfying the demand. A survey of the use of computers by official statisticians today compared to even 15 years ago would show a vast expansion of activity and a growing dependence on computers in order to achieve the objectives of official statistics.

28. No such comprehensive or detailed survey is given in the present document, but rather an attempt is made to describe below the present level of utilization of the various "information technologies", referred to in chapter I above and covering hardware and software.

29. At the national level in developed countries, the use of computers for statistical purposes continues to grow at a rapid pace. The phenomena noted in the introduction are especially apparent, in particular in the sense that the success of earlier efforts to computerize the processing of statistics has led, through a sort of "demonstration effect", to an increase in the demand for statistics.

30. Two important regional surveys of the use of computers for processing official statistics have been conducted. They differ significantly in character. The first is the set of national reports on the use of electronic data processing, submitted to the Working Group on Electronic Data Processing of the Conference of European Statisticians. 12/ The second is the review of national data processing capabilities in Africa, submitted to the Joint Conference of African Planners, Statisticians and Demographers. 13/

B. Trends and experience in official statistical offices
of developed countries

31. The summary account of the use of computers in Europe relates mostly to developed countries with a decade or more of experience in the use of medium-scale to large-scale computers. The reports (see para. 30 above) cover the United States of America, Australia and Canada as well as the European countries.

32. One impressive feature of the reports is the evidence they give of the recent rapid increase in the installation and use of data-base management systems. A longitudinal study of the reports covering the last decade indicates also a considerable increase in the power of the processors, a marked shift towards storage on disk (rather than on tape, although extensive tape processing facilities remain) and a definite growth in the use of generalized statistical software.

33. With respect to generalized software, it is encouraging to note to what extent a package developed in one national or international office is being used in others. Some principal examples are:

- (a) TPL (Table-Producing Language) - from the United States Bureau of Labor Statistics;
- (b) TAB68 - a tabulation package from the National Central Bureau of Statistics of Sweden;
- (c) AERO and CANEDIT - generalized editing packages developed in Hungary and Canada;
- (d) XTALLY and UNEDIT - generalized tabulation and editing systems developed by the United Nations;
- (e) RAPID - the data-base system developed by Statistics Canada;
- (f) COCENTS, CENTS - from the United States Bureau of the Census.

34. Ten or fifteen years ago, there was very little sharing. National offices were intent on constructing their own generalized or semi-generalized systems, sometimes with good reason in view of special hardware or statistical situations. However, genuine generalization is found to a higher degree in the more recently developed software because:

- (a) it is often developed with an eye to portability, being written in high-level languages which are available on many machines;
- (b) it deals with a wider range of common statistical problems;
- (c) through continual exchange of ideas and experiences, the statistical offices have adopted similar approaches to data processing.

35. There is no intention to suggest that these advances have come about easily. Davies 14/ notes that some countries had begun with the generalized systems that happened to be available on the machine in use but subsequently abandoned the effort because of the inefficiency of the method. He notes also that in other countries, several agencies have undertaken unco-ordinated development of software, with a consequential proliferation of data structures.

36. It could be said also that, as data processing in official statistical offices matured and became more sophisticated, it has come to be realized that problems are not unique and that a solution may well be found in one office for the problem in another.

37. As regards the hardware, perhaps the most significant new trends are those observable in output. First, there has been a rapid growth in the use of photocomposition facilities. For producing census publications, for example, photocomposition has been reported as being widely used in Finland, Canada, the United States and Japan. 15/ The same study indicates that about a third of the countries have built up sizeable libraries of magnetic tapes and microfiches of census results as an alternative to publications.

38. One innovation of great interest to statisticians is the ability to disseminate data down to the household and enterprise level. For example, the Central Bureau of Statistics of the Netherlands has reported on the use of the Viditel system to transmit data through the normal telephone network. 16/ The hardware of this system consists of a special computer with a capacity of 180,000 pages (each page being 40 lines of 24 characters). The user needs a colour television set with built-in decoder, a modem connection to the telephone system and a simple keyboard. The user may then gain access to and view the data by simple "tree-search" through a subject classification, through an alphabetic list of information suppliers or through a keyboard system. The results of this experiment are not yet complete, but future possibilities include two-way traffic. Similar experiments are also being carried out in other countries.

C. Trends and experience in official statistical offices of developing countries

39. The survey in Africa (see para. 30 above) illustrates recent experience in developing countries. Since the coverage was incomplete, the findings are indicative only. Nevertheless, some interesting observations can be made.

40. Although the use of computers for statistical purposes expanded considerably during the 1970s, more than half of the computers installed in Africa during that time are relatively small (32K to 128K bytes of primary storage, for example). If one examines the cycle-time of these machines, they rate from 56 nanoseconds to 11.5 microseconds.

41. By comparison, the original machines installed in the developed countries in the late 1950s or early 1960s may have had greater primary memory capacity (256K to 1024K bytes) but much slower processors (5 microseconds up to milliseconds). What

is noteworthy is that even in developing countries statistical offices are installing their own computers (for example, the National Statistical Office of Malawi recently acquired a powerful 64K micro-computer). Smaller statistical units in government departments may even acquire their own machines for specialized use.

42. Evidently, therefore, it was of great advantage to begin data processing in the 1970s rather than in the 1960s. The later computers were more sophisticated, faster and more powerful; they were served by peripheral devices that were faster and more reliable and produced better quality output; and more generalized, powerful and easy-to-use software was available. All these improved facilities were considerably cheaper in the 1970s than the inferior ones of the previous decade. The developing country initiating statistical computing in the 1970s certainly had the advantage of "jumping in down-stream" in the development of computers.

43. Powerful generalized statistical software packages have been developed by many national and international offices for use on large machines. Most of them could be used only on a large machine. However, software of comparable usefulness is becoming available on smaller machines. The UNEDIT and XTALLY packages developed by the United Nations 17/ are examples of statistical software for use on small machines. Boag reports some very sophisticated work being done on a microcomputer installation. 18/

44. Other trends in software and applications in statistical offices of developing countries include the wide use of generalized programs and the introduction of integrated processing through the use of expanded disk capacity.

D. Trends and experience in international statistical offices

45. The statistical offices of the international agencies have been using computing facilities for 20 years, the United Nations Statistical Office in New York being a pioneer in this respect. Today, the major installations are the New York Computing Service and the International Computing Centre at Geneva. These are linked by satellite, a link which has not yet been widely used for statistical applications.

46. The exchange of data in machine-readable form between the agencies has expanded considerably, and data and programs have also been transmitted by satellite experimentally. Discussion has begun among the agencies on the question of installing a network linking their various computing facilities. 19/

47. The agencies have been exchanging experience and information on both the development and use of data-base management systems. There has been some experience in the portability and sharing of such systems in that UNSIS (United Nations Statistical Information System) is installed both in New York and Geneva.

48. The trends and developments in international statistical offices are in general rather similar to those in the national offices of developed countries. In general, however, the quantity of data is relatively small. On the other hand, the

publication activity is relatively more important. Thus the move to photocomposition is very marked in international agencies. The United Nations, the Organisation for Economic Co-operation and Development and the International Monetary Fund produce almost all publications by this technique, and each has a small in-house facility. The other agencies are also introducing photocomposition at an accelerating pace.

49. At the moment, the use of meta data bases in national and international statistical offices is still fairly limited. However, as part of the co-operative software research and development project being carried out under the auspices of the Conference of European Statisticians, considerable attention is being given to the development of a meta information system. 20/ Various models exist in the many data-base management systems which provide directories, dictionaries, glossaries and indexes. One object of the research project is to devise a meta system especially designed for official statistics. The international agencies have had discussions on the adoption of standards, with the creation of indexes to the joint store of data being of special concern.

E. Trends and experience outside of official statistical systems

50. In both national and international statistical offices, the central processing units of the computers tend to be large and the processing tends to be centralized. However, a trend has started in other fields of statistical processing, primarily in business and in academic circles, towards the use of minicomputers or even microcomputers in the processing of statistical data.

51. The trend is significant because powerful tools for manipulating statistical data are now becoming more widely available to smaller organizations, developing countries, individuals etc. This development affects economic considerations involved in computerization and even the structure and distribution of work. The trend is elaborated in chapter III below.

52. As the statistical data bases have grown larger and more complex, and as statistical information systems have become more sophisticated, attention has been given to "meta data" or data about data. The meta data relate to the information system itself; they indicate what data exist, how they are coded and classified, where they are stored etc. - perhaps even what software is to be used for gaining access to and manipulating the data.

53. Finally, statistical packages for a variety of purposes have become much more widely available. There is a genuine problem of selecting the right set of packages, developed in commercial enterprises, research and university departments and government departments. There have been several attempts at comparative evaluation, notably by Ivor Francis. 21/

III. FUTURE POSSIBILITIES FOR EXPLOITING INFORMATION
TECHNOLOGY IN A STATISTICAL ENVIRONMENT

A. Statistical networks

54. It has been suggested that an ultimate world statistical system will be based on a global network, with participation by both national statistical offices and international agencies. 22/ A statistical network would make it possible to use telecommunication systems for transmitting statistical information.

1. National networks

55. A network is an interconnection of several computer systems and/or terminals with communications facilities. The participants in the network share the resources - the hardware, software and data - and distributed processing and storage can be applied. The use of such a network makes it possible to gain access to special software for some particular analysis, computer processors with special speed or size or certain data - none of which may be available in the user's own office. The primary objectives of such an approach are to reduce costs, improve the statistical service and share resources. The potential impact on national and international statistical agencies is enormous. Department (or region) X would not need to receive and store the data of department (or region) Y, because access to the data can be obtained when needed.

56. This notion, extended to its limit, might sound as if its application would drastically alter national statistical systems, and yet it is only an extension of some national systems already being established. The figure in annex I of the present document illustrates the main components of a national network. The national system may be operated as an independent system or part of a general system.

57. An associated innovation will be the extensive use of administrative records - a trend already started and noted above. As this trend continues, there will take shape at the national level an information system with a "finger on the pulse" of national economic and social development. The statistical information system can then be linked to the other systems in its environment - the respondents, the users and the "target" system being measured - that is, the socio-economic life of the country concerned. 23/

58. There is no intention to suggest that the monitoring will be "real time"; rather, it will be "real enough" for national planning and decision-making. The important consideration is that, in principle at least, as soon as the data are available to the original national authorities, they will be accessible to all national organizations.

2. International networks

59. Just as with some national networks, the communications facilities have already grown to the point of being able to support an international network. The objectives of such an international network would be the same as for a national system - to reduce costs, improve the service and share resources.

60. The lead in this field has been taken by the European Communities, following the lines established by the European Council at Dublin in 1979. The Council placed special emphasis on the development of information technologies in Europe, in particular: (a) creation of a homogeneous market of telematics equipment and services; and (b) creation of the European information industry.

61. The Statistical Office of the European Communities (EUROSTAT) has begun on-line dissemination of its data bases on EURONET - the communications network of the European Communities. At present, the dissemination takes place through a commercial agency, since the Computer Centre of the Commission of the European Communities is not yet linked to the network (although this link is planned). As direct links to EURONET become available, EUROSTAT intends to use them for disseminating data bases to privileged users and eventually to establish general communication links to one or more of the national statistical offices.

62. Among the international statistical agencies, there has been as yet little activity in linking computers across communication lines. The United Nations, however, has established a satellite link between New York and Geneva that has been used experimentally for the transmission of data and programs. The diagram in annex I shows the existing communications facilities of the United Nations.

63. The business community already has several information systems in place, supported by appropriate communications facilities. It is not unusual today for a company to transmit data from one country to a branch in another and for the latter to do the necessary computing and transmit the results back. This is done to take advantage of special hardware or cost factors. The communications and network technology is sufficiently far advanced to support this kind of activity, and a global statistical network is already technologically feasible.

64. The implications of this development, which many see as inevitable, at least in an important group of countries, will be double-edged. There will be great improvements in the statistical systems of participating countries, but also heavy investments will have to be made in planning, co-ordinating and maintaining the national monitoring and international network.

65. The notion of on-line statistical data bases has been taken up in a very practical sense by the international statistical agencies, and several have operational systems, generally for internal purposes only. However, some of their data are available through international service bureaux on commercial terms. Some of the agencies have begun discussing the possibility of making their data available directly on-line to national offices and other users. One problem foreseen is how to establish appropriate prices for on-line usage and it has been proposed that there should be discussion and co-ordination of a common policy. 24/

66. On the basis of current trends in co-operation between countries and between international agencies, one could postulate a global statistical network. In fact, the subject has already been discussed by the Sub-Committee on Statistical Activities of the Administrative Committee on Co-ordination (ACC). 25/ Such a network could be built up modularly, linking users and respondents to national and subnational offices. The national offices in turn could be linked to United Nations regional offices and international statistical agencies. The agencies themselves could be linked via satellite communication or terrestrial lines. The diagram in annex II shows a possible network, building on the existing United Nations facilities, which should be integrated with the statistical networks of other organizations.

67. The United Nations is currently negotiating with two organizations on facilities allowing broad-band communication by all United Nations ground stations. The success of these negotiations would hold out the prospect of great potential for international statistical information systems. Coupled with the network developments already under way at EUROSTAT and elsewhere, this expansion of communications may mean that the beginning of a truly international statistical information system may be no more than five years off.

68. One problem related to on-line systems and networks for which no appropriate system seems yet to have been developed is that of meta data, including in particular foot-notes. In official statistics, and especially at the international level, there is a need to store explanatory text along with the data and it must be possible to retrieve and publish the text automatically when the data are retrieved. This requirement has been partly met in some publication systems, for example those of the International Monetary Fund and the United Nations Statistical Office. 26/ 27/ However, it is of critical importance, when allowing "outsiders" access to an agency's data base (either national or international), to ensure that the explanatory text appears on the screen along with the data. In fact, it has been proposed that the user should receive the foot-notes and other supporting text automatically. 28/

69. In short, a great deal of work needs to be done in order to build a global network with its universal access to up-to-date (or "real enough") statistical data. The consequence would be that the statistical system will be in a position to offer vastly improved statistical services.

B. Impact on economic and social development

70. The most important feature of the future innovations mentioned above will be the direct linking of users, respondents and statistical offices. At the national level, the ability of the statistical service to provide up-to-date statistics on the national economic and social status will be greatly enhanced.

71. For purposes of national planning and decision-making, the statistical information system should produce statistics corresponding to three objectives: they should be timely, comprehensive and measure the entities in which the authorities are interested.

72. The linking of the official information system directly to the sources (official administrative records, enterprise computer systems, individual homes) should yield much more timely data. Provisional figures could be "real enough" for a close monitoring of economic activity, and the extension of the scope of the network will improve comprehensiveness.

73. It is important to note, however, that information technology is not the panacea for all statistical problems. The third objective, namely that the statistics should measure the entities appropriate to the needs of the planners and decision-makers, will not be directly influenced by advances in computing technology. In an indirect sense, by improving feedback and creating opportunities for flexible survey design, advanced information technology will assist in the formulation of appropriate statistical objectives, but the determination of the objectives will still be a non-computerized activity.

C. Impact on the production of statistics

74. One obvious effect of the use of the data-base technology is the greater demand for careful control, editing and correction of input. On the one hand, data-base management systems are not very tolerant of errors in the data and, on the other, the extension of on-line access to a host of users makes it imperative that the contents of the data base should be "clean". The design of forms for data collection must, therefore, take into account the fact that the form also becomes the input document to the computing process. This consideration will be of prime importance. Technological innovations such as optical character recognition and mark-sensing will be married to a conventional design of forms to meet the demand for reducing human effort and error.

75. Of the many options available for ensuring the accuracy of the data in machine-readable form, the idea of correcting them at a point as close to the source as possible seems to be the most practical. "Intelligent" terminals and data-entry devices can be used to check the data as they are being keyed and for prompting the operator to check material. This verification of data will become very important.

76. When computerized processing became possible in developing countries, the statisticians were taken somewhat by surprise and they had to rely very heavily on data processing experts. As a result of the application of the new technology, data collection and processing became unduly separated and hence organizational difficulties arose. The difficulties have gradually been alleviated by training and orientation. With the pending introduction of data bases and on-line systems, it is especially important for the professional statisticians to be involved in design and implementation.

77. The role of the data-base administrator will also become one of growing importance in the statistical office of the future. As data are integrated and stored in the data base, control over them will become more centralized. The data-base administrator will ultimately be responsible for determining which data are entered, who has access, how and where the data are stored, which data are

protected, the degree of protection (within the statistical office, the country, the region etc.), which data will be more accessible and so on.

78. Monitoring of data-base usage and dynamic restructuring will be especially important. Compared to a typical commercial data base, the official statistical data base grows very rapidly. Demands can also change quickly. Therefore, restructuring is of great strategic importance and must be carried out at regular intervals.

79. It will be necessary to endow the data base administrator with the authority, human resources and tools necessary for performing the task effectively and efficiently.

80. The data-base administrator will need, for example, software for interpreting the user's descriptions of the data proposed for input. There will also be a need for an extensive system of cataloguing. The catalogues need to contain not only information about the data stored but also information about the software that can be used, a cumulative record of usage, information about users and perhaps about sources as well.

81. Information about sources, users and type of on-line activity will be of increasing importance as the statistical office participates in network systems.

82. Since the tasks of the data-base administrator are largely technical, the position would normally be filled by a computer specialist. However, data stored in a data base must be carefully planned, co-ordinated, standardized, classified and described. Such a task involving statistical data is usually outside the competence of computer specialists and can only be undertaken by a statistician. Furthermore, statistical data within a data base must be made compatible for multiple information purposes which, in turn, necessitates the co-ordination and management of the data and provides organization-wide access to them. This requirement emphasizes the statistical aspect of the task. The creation of a data base and a data dictionary will make it possible to streamline the design of survey systems and to achieve the long-sought objectives of data integration. But the benefits will not materialize without the participation of statisticians as data co-ordinators to organize the service-wide co-ordination and management of data. This is essentially a departure from the traditional functions of statistical services but since it was first suggested, 29/ it has been adopted by several statistical services. The function includes at least the development of data definitions, standards, classifications and procedures and documentation of information about the data.

83. The major elements of standardization include the identification of the physical attributes of the data and ensuring that the data are consistent as regards similar characteristics and attributes. Classifications are required in order to relate data to a uniformly designated structure so as to ensure that they will be consistent and compatible for all cross-referencing. A further advantage of the standardization and classification of data is that reclassification could be performed automatically by the use of algorithms.

84. New data to be added to the data base need to be adequately described for inclusion in the data dictionary. Such documentation will ensure that uniform descriptions of data can be produced, that data can be easily referenced and that the attributes of data can automatically be included in application programs. The documentation will have to be reviewed periodically so that the data dictionary will not include any duplication and so that redundant elements may be deleted.

85. Most statistical offices, at the national and international level, have been organized along subject-matter lines. The personnel, the publications, the data and the computer systems have generally been separated according to major branches of statistics. These branches have generally functioned independently of each other. The computer systems which were established simply reflected the distinctions and demarcations.

86. With the advent of the integrated data base, this practice must change. Major modifications in procedures, computer systems and even organization are necessary. The integration of data requires the breaking down of the traditional boundaries. The concept of an integrated data base implies a sharing of access to and use of data. This, in turn, demands common systems for editing, imputation and correction so that the data are of even, consistent quality. It demands also a much greater level of control and co-ordination than has traditionally been the case.

87. The introduction of national networks will greatly increase the possibility of using administrative records for statistical purposes. Statistical offices could conceivably gain access to administrative files of various government departments in order to extract statistics as required. Given the necessary co-operation, the official statistician could even tap the computerized files of enterprises. This possibility will grow as the world continues to become more computerized.

88. At the national level, far-reaching developments in communications technology have taken place in the last decade in developed countries. The beginnings of a similar explosion are seen in many developing countries. At the same time there is the beginning of decentralized computing and data-entry facilities. The joint effect will be the geographical decentralization of storage, processing and access. Statistical data - and administrative records - can be maintained at, or close to, the sources.

89. During the past 25 years or so, the introduction of computers into government statistical services has meant not only that statistical management had to deal with the technical problems of a new technology but also had to adjust and adapt to the impact of the technology. The technical problems of exploiting the hardware of successive generations of computers have come and gone. The problems of adjusting and adapting the statistical services to the impact of computers are slowly coming to be understood and are either partially solved or ignored. But as the use of computers in statistical services has evolved from simple batch processing to sophisticated statistical information systems, the problems of how to utilize them fully to the benefit of the population as a whole have hardly begun to be considered. For example, an adequate flow of data and information may already exist at the local or rural level, but the information does not flow spontaneously to the level of urban-based statisticians and planners. Nor do statistical

information or the plans conceived by government find their way spontaneously to the local or rural level. The information system required to support statistical activities and government action in order to achieve development objectives needs to be effective enough to bridge gaps between people who would not normally communicate with each other.

90. The major implications of these developments for the work of the statistician will be, on the one hand, a demand for very stringent control and standardization and, on the other, the creation of innumerable new data and manipulation possibilities.

91. Information analysis is seen as an emerging, disciplined methodology for generating formal descriptions of the information content of data bases. Given proper training, the information analysts in a statistical office can take steps towards the management of information resources. 30/

92. Finally, the availability of data, text and documentation on an easily accessible network opens up the possibility of "electronic commuting" for the staff of statistical offices. The statistician of the future could conceivably do most of the work at home by using terminals and a simple telephone connection.

93. The beginnings of most of what has been described above are already visible. What is being predicted is a dramatic extension of existing trends throughout statistical services and across national boundaries. A global statistical information system, for example, is technologically feasible.

D. Impact on the users of statistics

94. The users of official statistics in the future will have at their disposal not only an increased fund of statistics, more up to date, comprehensive and reliable than before, but also statistics in a variety of media, including on-line access to network systems. One immediate result for the user will be that the way will be open to a wider, richer analysis of statistics. This, in turn, may foster a growth in demand for even more statistics.

95. The official statistical office of the future may well operate for the user as a kind of statistical service bureau, giving information on the availability and use of statistics and providing tools for access to and the extraction, manipulation and analysis of the data - in addition to providing the data themselves, or access to them. The official statistician may even provide training in the use of the data-base system and associated software tools. The training could take the form of computer-assisted instruction (CAI) to the user seated at a terminal.

96. For the users connected via terminals, or subscribing to a cable TV channel with an official statistical service, it would be possible to broadcast promotional information announcing new data or new services and facilities.

E. Impact on respondents

97. Information is probably the most easily obtained and shared resource for development (compared, say, to land, labour and capital). Nevertheless, it is recognized that its contribution to development is growing in importance, as evidenced by the investment taking place in national and international infrastructures and systems. It is important not to overlook the impact of the developments on the sources of statistics, that is, the respondents, whether government or semi-government organs, private enterprises or individuals.

98. Without adequate planning the inevitable effect of an extensive network approach on the respondent would be to create burden. Proper use of the improved technology could largely ease the burden. For example, the use of administrative records, computerised and accessed on-line by the official statistician, could lighten the response load on government departments or private enterprises.

99. Confidentiality is likely to become of serious concern in such a system. Attention must be given to formulating careful controls in order to avoid deliberate or accidental access to confidential information - either in the data base of the respondent or in that of the official statistician. One consequence of the new technology will be that the legislation concerning confidentiality may have to be reviewed. There is an additional aspect requiring attention: communications networks may be tapped by unauthorized persons and technical measures such as cryptography may have to be applied so as to safeguard security and confidentiality.

100. The linking of the official statistical system to the respondent's data bases will open up possibilities of flexibility in conducting ad hoc surveys. Although many attempts at linkage have foundered on the lack of standardization, greater attention to the co-ordinated use of standard codes and classifications will lead to the linking possibilities.

101. As far as individuals are concerned, the use of television sets specially equipped to receive and store data transmitted over cable or satellite links may extend the possibility of direct inquiry into the private home.

IV. SUMMARY

102. The possibilities referred to above will not materialize easily or without a considerable investment in:

(a) Co-operative efforts to develop common standards, codes, classifications and definitions;

(b) Extensive and intensive co-operative methodological work in system design and development;

(c) Extensive technical co-operation efforts to improve national statistical systems in developing countries and to promote and accelerate training in data-base technology;

(d) A systematic sharing of solutions to statistical computing problems and, where possible, the exchange of software of mutual utility.

103. The fact that the Statistical Commission chose to discuss the impact of information technology now is evidence of a realization that it is time to begin co-operation in this field. Various national agencies have advanced plans for data bases and networks, on-line access, data transmission etc. Without conscious effort, system designs can become unco-ordinated and incompatible. As mentioned in paragraph 49 above, under the auspices of the Conference of European Statisticians, a co-operative research and development project has been initiated in the European region. The Statistical Office of the European Communities has also begun using the EURONET system. The results of these efforts should be watched closely by all national and international statistical agencies.

V. POINTS FOR DISCUSSION

104. The Commission may wish, in view of the importance of the foregoing considerations, to discuss the following:

(a) The relevance of the new technologies to improvements in statistical systems at the national and international levels;

(b) The possibility that, unless statistical agencies assume the responsibility for a collection and dissemination network, organizations without the requisite competence may do so;

(c) The steps necessary for designing and planning a world statistical network.

Notes

- 1/ Official Records of the Economic and Social Council, Thirtieth Session, Supplement No. 12 (E/3375), para. 158 (4) (m).
- 2/ Ibid., 1981, Supplement No. 2 (E/1981/2), paras. 267-268.
- 3/ "Computing in national statistical services beyond 1980" (CES/WG.9/141).
- 4/ "Report of the Study Group on the Co-ordination of Government Information Systems on its second session" (ADM/CGIS(2)/5).
- 5/ Ibid., para. 43.
- 6/ "Seminar on Statistical Data Collection and Processing Systems under New Conditions; Report of the Seminar held in Moscow, 21-25 September 1981" (CES/SEM.14/3).
- 7/ "The dissemination of official statistics-technology", by Tore Dalenius, paper presented to the forty-third session of the International Statistical Institute, Buenos Aires, 1981.
- 8/ "System analysis and design applied to statistical systems" by Joan C. and Svein Nordbotten, paper presented to the forty-third session of the International Statistical Institute, Buenos Aires, 1981.
- 9/ Social Indicators III, United States Bureau of the Census, Washington, D.C., December 1980.
- 10/ "Understanding the computer and its impact on society", Datapro Directory of Small Computers, Datapro Research Corporation, Delran, New Jersey 08075, USA, 1980. SC10-100-101 - 104.
- 11/ "Use of administrative records for statistical purposes", submitted to the present session of the Statistical Commission (E/CN.3/1983/2).
- 12/ "National reports on the use of electronic data processing for statistical purposes", Conference of European Statisticians Working Party on Electronic Data Processing (CES/WP.9/(various numbers)).
- 13/ "Review of national data processing capabilities in Africa", (ST/ECA/PSD.2/15).
- 14/ "Data base management in national statistical agencies" by Barrie M. Davies, paper presented to the International Statistical Institute/International Association for Statistical Computing meeting on statistical data bases, Manila, 1979.
- 15/ Provisional results of the United Nations Statistical Office study on population census costs and staffing requirements, unpublished.

16/ "Experiences with a new form of statistical dissemination", prepared by the Netherlands Central Bureau of Statistics for the Seminar on Statistical Data Collection and Processing Systems under New Conditions, Moscow, 21-25 September 1981 (CES/SEM.14/R.3).

17/ I. Francis, Statistical Software - A Comparative Review. New York: North-Holland, 1981, 542 pp.

18/ "Microcomputers and the statisticians" by I. Boag, paper presented to the forty-third session of the International Statistical Institute, Buenos Aires, 1981.

19/ "Outline of a global statistical network", paper presented to the ACC Sub-Committee on Statistical Activities, sixteenth session, Vienna, 1982 (SA/1982/8/Add.1).

20/ "Progress report on the activities of the ECE/UNDP statistical computing project in 1981", Conference of European Statisticians, thirtieth plenary session, Geneva, 1982 (CES/470).

21/ Statistical Software - A Comparative Review, op. cit.

22/ "Outline of a global statistical network", op. cit.

23/ "System analysis and design applied to statistical systems", op. cit.

24/ "Report of the ACC Sub-Committee on Statistical Activities on its fifteenth session", Washington, 1981 (ACC/1981/26), para. 26.

25/ "Outline of a global statistical network", op. cit.

26/ A Guide to the Data Fund, International Monetary Fund, Washington, D.C., to be published in 1982.

27/ "An approach to computerized publications", paper presented to the sixteenth session of the Working Party on Electronic Data Processing, Conference of European Statisticians, Geneva, May 1979.

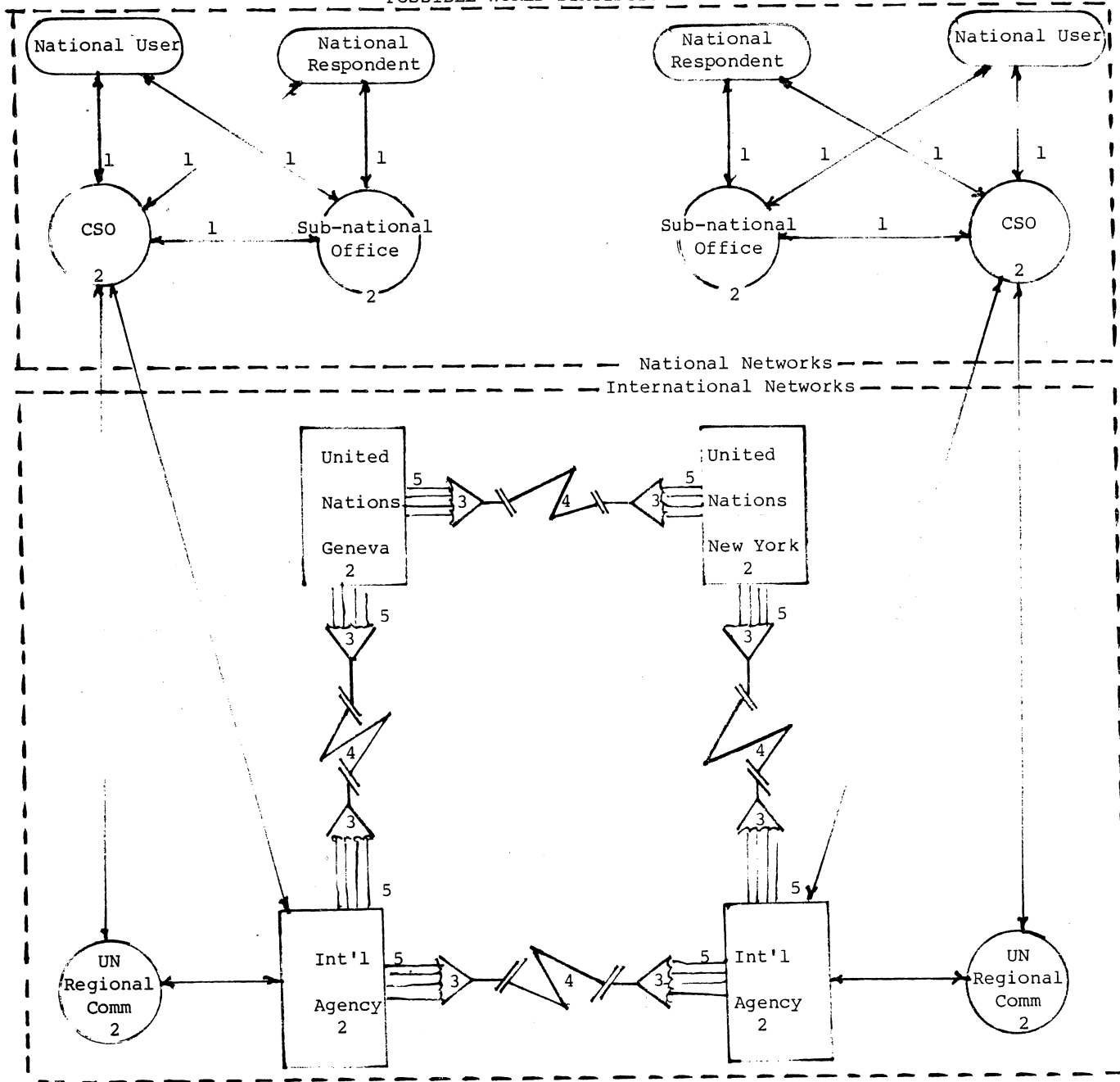
28/ "Outline of a global statistical network", op. cit.

29/ "Computing in national statistical services beyond 1980", op. cit.

30/ Information Resource Management: Concept and Cases, Forest W. Horton Jr., Association for Systems Management, Library of Congress Catalogue No. 79-53322.

Annex I

POSSIBLE WORLD STATISTICAL NETWORK



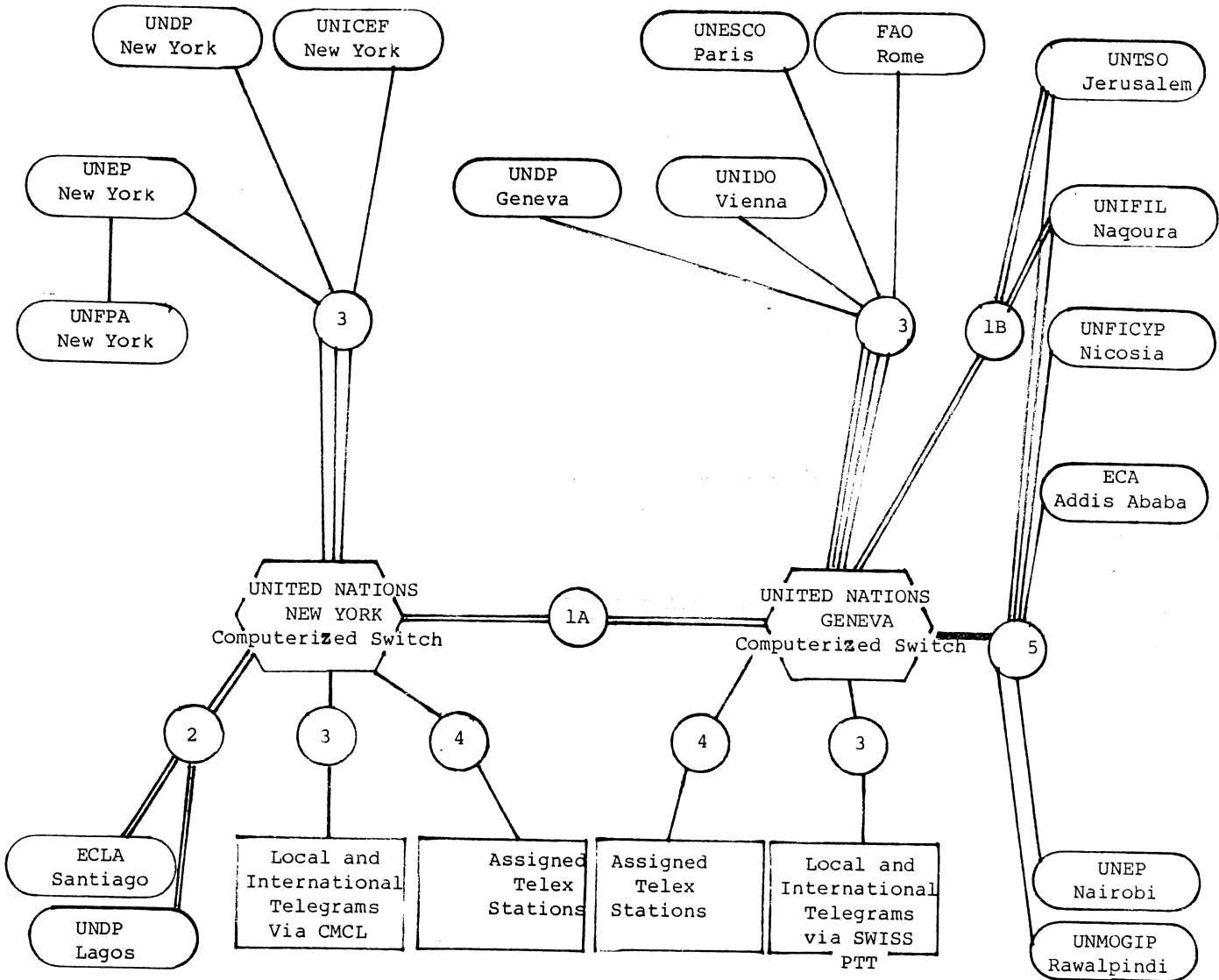
Legend

- 1. Dial-up Telephone Lines
- 2. Computing Facilities
- 3. Statistical Multiplexer
- 4. Satellite Link
- 5. Bi-sync communications link
- CSO. central statistical office

/...

Annex II

EXISTING UNITED NATIONS COMMUNICATION SYSTEM



- Legend:
- 1 A AVD/TTY circuit via INTELSAT Satellite
 - 1 B VOICE/TTY circuit via SYMPHONIE Satellite
 - 2 TTY Satellite leased circuits
 - 3 Dedicated lines
 - 4 Telex circuits
 - 5 H.F. Radio circuits owned and operated by United Nations

ECA Economic Commission for Africa

ECLA Economic Commission for Latin America

FAO Food and Agriculture Organization of the United Nations

UNDP United Nations Development Programme

UNEP United Nations Environment Programme

UNESCO United Nations Educational, Scientific and Cultural Organization

UNFICYP United Nations Force in Cyprus

UNFPA United Nations Fund for Population Activities

UNICEF United Nations Children's Fund

UNIDO United Nations Industrial Development Organization

UNIFIL United Nations Interim Force in Lebanon

UNMOGIP United Nations Military Observer Group in India and Pakistan

UNTSO United Nations Truce Supervision Organization in Palestine
