

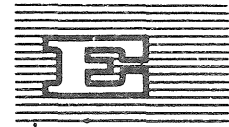
UNITED NATIONS
ECONOMIC
AND
SOCIAL COUNCIL



Distr.
GENERAL

E/CN.3/380
9 January 1968

ORIGINAL: ENGLISH



STATISTICAL COMMISSION
Fifteenth session
Item 8 (e) of the provisional agenda

DEMOGRAPHIC AND HOUSING STATISTICS

PROGRESS REPORT ON COMPUTERIZATION OF DEMOGRAPHIC STATISTICS

Report of the Secretary-General

TABLE OF CONTENTS

	<u>Paragraphs</u>
A. Data bank - macro level.....	6 - 21
1. Backlog operation.....	9 - 16
2. Current data procedures.....	17 - 21
B. Data bank - micro level.....	22 - 24
C. Generalized computer programme	25 - 34
1. Objectives.....	26
2. Progress.....	27 - 34
D. Dissemination of demographic statistics.....	35 - 38
E. Bibliography.....	39 - 40
1. Objectives.....	39
2. Progress.....	40
ANNEX. <u>List of demographic statistics series collected on</u> <u>Demographic Yearbook questionnaires.</u>	

1. This document serves to present to the Commission information on the present status of work initiated at its thirteenth session^{1/} on the application of electronic data processing to the compilation and tabulation of international demographic statistics.

2. At its thirteenth session, the Commission had before it a proposal for the establishment, in the Secretariat, of a "data bank" on magnetic tape, of national population census results. The proposal for consideration at that session was to collect a random sample of punch cards or equivalent statistical processing units from the 182 national censuses taken around 1960. It was proposed that, in due course, corresponding data would be assembled from the 1970 round of population censuses, and eventually compilation of vital statistics would also be computerized in order to facilitate the computing of population projections, life tables as well as rates and ratios required for demographic research. These data would also form an experimental data bank which could be used for making selected distributions required for research purposes.

3. The Commission felt that the proposal, as originally described, was likely to involve countries in problems of a legal, financial and technical nature. Members agreed, however, that more information was necessary before a considered judgment could be made and they therefore "requested the Secretary-General (1) to continue to study the collection, tabulation and analysis of population census statistics at the international level in the light of possible applications of electronic data processing, and (2) to prepare a report on the scope of the project for a centre for population data and the procedures necessary to implement it, for consideration by the Statistical Commission".^{2/}

1/ Official Records of the Economic and Social Council, Thirty-ninth session, Supplement No.13, (E/4045), para. 163.

2/ Ibid., para. 163.

4. In response to the Commission's request, as well as to clarify the scope of the work, a project was drawn up with the following objectives:

(a) To establish in the United Nations International Computing Centre a computer-based system for storage, adjustment, retrieval and analysis of demographic statistics - by means of a data bank of international demographic statistics, recorded on magnetic tapes;

(b) To develop a generalized computer programme for prompt tabulation of national demographic data according to both national and international standards;

(c) To devise ways to disseminate stored statistics currently;

(d) To establish a bibliography of official sources of demographic statistics on magnetic tape.

5. This report describes the progress made in each aspect of the computerization project beginning with the "data bank".

A. Data bank - macro level

6. Since both the Statistical Commission and the Population Commission had had reservations about the possibility of beginning with a data bank at the micro- or raw-data level, it was decided to proceed on the level of aggregated statistics, that is, to computerize data in the tabulated form in which they are currently collected from Governments. This meant beginning with the demographic statistics (population, vital and international migration statistics) which have been assembled since 1947 by means of annual questionnaires covering population from censuses and surveys, vital statistics from registers and surveys and migration statistics in a variety of cross-classifications suggested by the Statistical Commission at its second session when it established the contents of the United Nations Demographic Yearbook.

7. These series have been transcribed to "data cards" and filed by subject matter in the manually-operated data bank from which manuscript for publications and studies is currently prepared. Data collected by means of the Monthly

Bulletin of Statistics questionnaires, as well as those obtained from official publications and by correspondence, are routinely added to this basic bank of demographic statistics. The names of the 82 series which currently comprise the bank are set forth in the annex together with an estimate of the number of punch cards which would be needed to record the data.

8. The basic facilities required for implementing the computerization project were made available with the coming into operation, on 15 November 1965, of the United Nations International Computing Centre presently equipped with an IBM 7044 computer and an IBM 360 model 30. The Centre is designed to provide computing services to all units of the United Nations (statistical, accounting, financial and administrative) and, upon request and against payment of costs, to provide services to specialized agencies and other United Nations bodies, to Governments and private institutions. A description of the Centre's current capacities is found in E/CN.3/373, paras.88-91.

1. Backlog operation

a. Source documents

9. Tear sheets. For ease of operation, it was decided that the principal source documents for computerizing the backlog of demographic statistics should be the 17 volumes of the Demographic Yearbook, that is, tear sheets from the 1948-1965 issues, rather than the "data cards" described in para.7. About 80 subjects are included in these 17 volumes, some up-dated in each issue - others every five years.

10. In accordance with the rotation schedule for the feature topic in the Demographic Yearbook, the subject which has been examined in detail in the 1966 issue is "Trends in Mortality Statistics". It was therefore decided to attempt to use the computer to calculate the mortality rates specific for age and sex and, with this in view, priority for coding and punching was given to (a) population by age and sex and (b) deaths by age and sex. Work was initiated on devising codes for these subjects and punching of the population data began

January 1967.

b. Coding and punching

11. By 1 January 1968, backlog data from eleven tables^{3/} of the eighty-two shown in the annex had been coded; of these, data from seven (B-06, B-07, B-14, B-17, E-04, G-02 and G-03) had been punched on 62, 000 cards. Still to be punched are data from the remaining four coded tables, which should require some 30, 000 cards.

12. Backlog tables still to be coded as of January 1968 number seventy-one and these would require roughly an estimated 340, 000 cards.

13. It must be emphasized that the above estimates are a minimum, because they are based on published tables as they have appeared in the Demographic Yearbook. Not taken into account is the fact that, because of space limitations, many tables in the Demographic Yearbook show classification schemes which are not as detailed as those in the original questionnaires. For example, published data on deaths by age and cause show age consolidated into eight categories, while the series, as collected, is in 23 age groups. Likewise, foetal deaths by gestational age are published in 8 age groups (in weeks) while the original questionnaire shows 25 single weeks of gestational age, beginning at 17, plus a "less than 17 week" group. Hence, the number of cards estimated from the printed tables of the Demographic Yearbook, are a minimum. Also to be added

-
- ^{3/} B-06 - Population by 5-year age groups and sex.
B-07 - Population by single-years of age and sex.
B-14 - Population by school attendance, age and sex.
B-17 - Population by economic activity status, age and sex.
E-01 - Live births (total).
E-04 - Live births by age of mother and sex of child.
G-01 - Deaths (total).
G-02 - Deaths by age and sex.
G-03 - Deaths by cause and sex.
G-04 - Deaths by cause, age and sex.
M-C2 - Marriages by age of bride and by age of groom.

to the estimate of work still to be done are the data received on the questionnaires for the 1966 Demographic Yearbook and the 1967 Demographic Yearbook, estimated to require 70, 000 cards. Together with the 1948-1965 backlog, these make a total of 500, 000 cards to be punched as of the end of 1967 in order to computerize the data bank of population and vital statistics in aggregated form. This does not include international migration statistics.

14. It is estimated that one coder can complete three tables corresponding to about 15, 000 cards per month. At that rate it would take one coder twenty-six months to code the population and vital statistics which were in the Statistical Office of the United Nations by the end of 1967. To complete this coding operation by the end of 1968 will require three coders, and steps are being taken to obtain the necessary resources. Transfer of the coded figures to punch cards will require around forty man-months and arrangements for this work are being explored.

c. Programming

15. Transfer of figures from punch cards to magnetic tape involves first checking procedures to establish internal consistency of distributions as well as intra-series compatibility. The time-table for this operation will depend on the priority given to this project in respect of computer time.

16. Programmes have already been written and successfully tested for population by age and sex and deaths by age and sex. Computation of age-sex specific mortality rates was completed by 1 July 1967. Since all manuscript for the 1966 Demographic Yearbook was already at the typesetter by 1 June 1967, it was too late to take advantage of computer processing for this table in that issue of the Demographic Yearbook, but, it is hoped that manuscript for the comparable table in the 1967 issue can be made by the computer, as well as for the table showing deaths by cause, age and sex.

2. Current data procedures

a. Source documents

17. (i) Questionnaires. Beginning with the 1967 Demographic Yearbook questionnaires, arrangements were made to process receipts immediately into the

computer for storage on magnetic tape.

18. A procedure has been set up whereby each questionnaire, upon receipt, is carefully scrutinized by area specialists for inconsistencies and discontinuities. When the questionnaire is "accredited", it is given a "batch-number" which identifies it as to year, type and serial number. A photo-copy of each accredited page of the questionnaire is then made for processing purposes. After delineation and coding, punching is done directly from these pages.

19. (ii) Tapes. At present, aggregated data collected are limited to those specified on the Demographic Yearbook questionnaires. Consideration may need to be given to collecting, instead, data recorded on magnetic tape at the national level. This method of collection could apply equally to population census results and vital statistics and it would automatically make available a great many more cross-classifications of data than are now collected, provided the data stored on tapes generated in the various countries could be accepted and stored by the United Nations computer for subsequent reorganization and retrieval. This, of course, would require that the data be expressed in standard units and identified by standard codes or, alternatively, be convertible to such standards.

20. If such a system could be developed, it could relieve the countries of the hardship of completing questionnaires annually and it might well be the most economical and rational means of meeting international needs. There are, for example, at least 8 major countries which took a census in 1965 and 17 which took one in 1966; some, if not all of these 25 censuses, will be tabulated on computers. It is conceivable therefore, that the results of these censuses, could be made available in tape form for international use and publication. This possibility will be actively explored in order that the method may be applied for the collection of the results of the large number of censuses expected to take place beginning around 1970.

b. Coding and punching

21. Current coding of questionnaires is proceeding according to schedule. During October, November and December 1967, a total of 151 out of a potential

157 batches of 1967 Demographic Yearbook questionnaires were coded. However, of these only 18 batches were punched. Unless this deficiency in punching can be eliminated, the plans envisioned in para.18 will not succeed. Efforts are being made to solve the difficulties.

B. Data bank - micro level

22. In response to the request of the Commission at its thirteenth session^{4/} for continued study, some thought has been given to the possibility of implementing, on a limited scale, the original proposal which was to set up a data bank of a sample of population statistics in the "raw" form of reports on individuals as recorded at the censuses. As noted in para.3, this proposal might raise questions of a legal nature in some countries, but it is obvious that these problems are not universal since several countries have already made available sample decks of population census schedules for demographic research. At least 17 countries have such samples available; among them are the United States (with "The One-in-A-Thousand Sample" of the 1960 population), fifteen countries in Latin America which have provided sample decks of punch cards from the 1960 censuses to the Centro Latinoamericano de Demografía (CELADE), the Philippines which supplied a similar deck to the Philippines Statistical Centre and Western Samoa. In addition, Japan has signified its willingness to collaborate in such an experiment with cards from the 1965 census.

23. The information contained in such sample decks would comprise substantially all of the characteristics of the sample of persons enumerated in the census in question, with the names and certain other identifying information having been eliminated at the source. Utilizing these reels of tape or punch cards,

^{4/} Official Records of the Economic and Social Council, Thirty-ninth session, Supplement No.13, (E/4045), para.163.

it is conceivable that tabulations for the analysis of fertility, internal migration, education, labour force etc., not now available from national tabulation programmes, could be made as required for the use of the United Nations Population Division and other authorized consumers in pursuance of their programme of studies. Alternatively, tabulations could be made on a standard basis to provide comparable data for regional studies. The latter activities would also provide a very useful methodological insight into the problems of standardizing definitions and classifications.

24. In view of the immediate availability of a relatively large number of sample decks, and the urgent need for special tabulations and for experimentation with standardization, it is proposed to collect such decks as may be available free of charge to the United Nations and to proceed with studies based on these data. Other countries could be encouraged to consider participating in this programme with samples from the 1970 censuses.

C. Generalized computer programme^{5/}

25. This part of the computerization project had its origin in the stress placed by the Commission at its thirteenth session on "the need for prompt tabulation and publication of results" of the 1970 censuses and the importance of internationally "comparable data for planning and evaluating international programmes....."^{6/}

1. Objectives

26. The objectives of the project would be to provide Governments, especially in developing countries where trained programmers are few, with a generalized

^{5/} "Programme" is used here in a broader sense than is usual in computer terminology.

^{6/} Official Records of the Economic and Social Council, Thirty-ninth session, Supplement No.13, (E/4045), paras.136-137.

or standard computer programme for processing the 1970 population census results. Such a programme might be part of a central library of computer programmes which could be supplied to Governments with limited resources for programming or those who plan to have their censuses processed at installations outside their own borders. Such programmes could cover analysis as well as tabulations.

2. Progress

27. The key concepts of a general tabulation programme have now been decided; they will be: "one record per destination", tables for recoding, and logic functions. "One record per destination" means that, if amounts in an input record have to be added to amounts in N different tables, N different records will be created, each carrying the identification of the table and the number of the place within the table. If the calculated destination is in the central memory, the amounts will be added there. Otherwise the amounts will be dumped on an output file (to be processed later) or part of the central memory will be dumped and space allocated there for the calculated destination. If a direct access storage device is available, the central memory will be used for the most frequent cases and the direct access storage device for the next most frequent, leaving a minimum to be dumped. Control-cards will tell the cases in the central memory from the cases in the direct access storage device.

28. The tables for recoding will tell, for each characteristic of an input record and for each of its corresponding output records, whether the input code shall remain unchanged or how it shall be changed.

29. The logic functions will contain the conditions for creation of new records.

30. The tables for recoding and the logic functions will be in control-cards. There will thus be a complete flexibility in both input and output coding. National and international tables can be produced simultaneously.

31. The above procedures have to be repeated, with new sets of control-cards, on each (geographical) level until all tables have been created. The tables for

recoding will make possible such coding that cannot be made until the processing has proceeded up to a certain level.

32. Consistency-tests must precede the tabulation. Only records with valid codes and valid combinations of codes will be accepted by the tabulating programme. Control-cards, containing all permissible codes, and logic functions, expressing the validity of relations between codes, will play a main role in a general testing programme. There are, however, cases where a complete testing before the first tabulation cannot conveniently be done; there must therefore be, besides the pretabulation tests, provisions for inter- and post-tabulation tests of the original input records. Errors in these later tests might affect tabulations already made. The tabulation programme will let records, originally accepted but later invalidated, up-date all summary records negatively at the same time as the corrections are up-dating them positively.

33. An individual record might be valid per se but none the less appear unacceptable when looked upon in its tabulated environment. This means that tables must be tested on each level and before they are aggregated to the next higher level. The general programmes must be able to signal frequencies or distributions that deviate significantly from expected values.

34. The Statistical Office expects to start the testing of a first version of a general tabulation programme before April 1968. The first version will be written in FORTRAN and, after debugging on the IBM 7077, compilations will be tried on computer of other models and from other manufacturers.

D. Dissemination of demographic statistics

35. The third part of the project envisages a revolution in dissemination of demographic statistics because computerization could provide speedier and more accurate methods of retrieval, reproduction and dissemination of data. At the moment, demographic statistics assembled in the Statistical Office are disseminated in periodic (annual, quarterly and monthly) publications, utilizing letter-press

from manuscript for the Demographic Yearbook and Monthly Bulletin of Statistics, and photo offset from fair copy for the quarterly Population and Vital Statistics Report. This is a slow and potentially inaccurate method of retrieval and reproduction. Up-dating of stored information, retrieval, reduction, adjustment, calculation and print-out by computer and reproduction by photo-offset, would constitute a major advance.

36. This aspect is being deferred pending completion of the backlog and current adjustment, storage and retrieval phases of the project described in paras. 6-24. It appears likely that this schedule will also coincide with availability of new devices which will make possible an improved manuscript for reproduction and dissemination. However, taking advantage of the experience gained in the computerization of external trade statistics, it is proposed that the intervening period should be utilized to examine the advisability of introducing more useful formats and more adequate means of dissemination such as a subscription service, as well as new publications, and conversely, the consolidation and/or elimination of existing publications.

37. Because of the volume of demographic statistics available annually from censuses, surveys and civil registers, the publication of these data was early put on a rotation basis in so far as detail was concerned. The cycle is approximately a 5-year one and this is quite adequate for population census data. However, it is not satisfactory for vital statistics, which become available in all types of cross-classifications annually.

38. It is conceivable, for example, that, when the backlog of demographic data are all on tape and current data are being processed routinely, the International Computing Centre could make available on request (ad hoc or subscription) and in the out-put form requested (i.e., cards, tape, print-out) any and all information in its possession, in any recombination or organization desired. The salient feature would be full access by participants to all

information made available by countries. The periodicity might be annual, monthly, or even shorter intervals. Such a system would be most advantageous; it would not, however, necessarily imply that publication of the Demographic Yearbook would cease. Rather, the Yearbook might be re-designed to serve as a reference book of final figures, while provisional data appeared in other forms. These ideas will be explored for future consideration.

E. Bibliography

1. Objectives

39. The fourth part of the computerization project is the related field of bibliography in which computerization will provide the means for advancement. The need for a "bibliography of primary source of national official demographic statistics" was recognized as far back as 1947 when the Population Commission recommended^{7/} that such a bibliography should be contained in the Demographic Yearbook. Accordingly, the Demographic Yearbook contained such a bibliography until the 1955 issue at which time matters of internal policy and shrinking space eliminated it. Despite requests by the Population Commission and the Statistical Commission, it has not proven feasible to publish a demographic bibliography since that time. With the advent of the computer, it becomes possible to consider the preparation of such a bibliography in the work programme. Its exact scope and detail has not yet been worked out but it is conceivable that it will include relatively detailed descriptions of content.

2. Progress

40. Entries for the bibliography are collected annually but nothing has yet been done regarding their computerization. It is estimated that for the period 1955-1964 there would be a minimum of 3, 000 entries. If these are indexed by subject matter as represented by the eighty-two distributions on which data are collected, the number of separate entries would amount to over 400,000.

^{7/} Official Records of the Economic and Social Council, Sixth session, Supplement No.4.

ANNEX

List of demographic statistics series collected on
Demographic Yearbook questionnaires

Table No.	Title	Estimated number of punch cards
<u>A. AREA</u>		
A01	Area by country	1,000
<u>B. POPULATION</u>		
B01	Total population by sex	5,000
B02	Population by major civil division	8,000
B03	" by urban-rural and sex	3,000
B04	" of capital cities and cities of 100,000 or more	14,000
B05	" by locality-size group and sex	3,000
B06	" by age (5-year) and sex (total/urban/rural)	15,000
B07	" by single years of age and sex	10,000
B08	" by marital status, age and sex	12,000
B09	" by country of birth and sex	16,000
B10	" by country of citizenship and sex	13,000
B11	" by ethnic composition and sex	7,000
B12	" by language and sex	5,000
B13	" by religion and sex	5,000
B14	" by school attendance, age and sex	2,000
B15	" by level of education, age and sex	8,000

Table No.	Title	Estimated number of punch cards
<u>B. POPULATION (continued)</u>		
B16	Population by literacy, age and sex	4,000
B17	" by economic activity status, age and sex	3,000
B18	Economically active population by industry, age and sex	6,000
B19	" by industry, status and sex	6,000
B20	" by occupation, age and sex	2,000
B21	" by occupation, status and sex	3,000
B22	" active population by status, age and sex	3,000
B23	" active women by marital status and age	700
B24	" inactive population by type of non-economic activity, age and sex	3,000
B25	Female population by age and size of family	3,000
B26	Female population by age and size of surviving family	1,000
B27	Population by type and size of household	1,000
<u>C. LOCALITY</u>		
C01	Number of localities by size	1,000
<u>D. HOUSEHOLD</u>		
D01	Number of households by type and size	1,000
<u>E. MORTALITY</u>		
E01	Live births (total)	2,000
E02	" by sex	2,000
E03	" by month of birth	1,000
E04	" by age of mother and sex of child (total, legitimate)	5,500
E05	" by age of mother (urban/rural)	500

Table No. Title Estimated number of punch cards

E. NATALITY (continued)

E06	Live births by age of father (total, legitimate) . . .	1,500
E07	" by birth order	1,000
E08	" by birth order and age of mother	12,000
E09	" by legitimacy status and sex	500
E10	" by type of birth	200
E11	" by occupation of father	100
E12	Legitimate live births by duration of marriage . . .	300

F. REPRCDUCTION RATES

F01	Gross and net reproduction rates	500
---------------	--	-----

G. GENERAL MCR TALI TY

G01	Deaths (total)	2,000
G02	" by age and sex	16,000
G03	" by cause and sex	16,000
G04	" by cause, age and sex	15,000
G05	" by cause and type of medical certification	500
G06	" by marital status, age and sex	2,000
G07	" by month of death	500
G08	" by occupation, age and sex	1,000

H. INFANT MCR TALI TY

H01	Infant deaths (total)	1,500
H02	" by age and sex	5,000
H03	" by month of death	200

Table No.	Title	Estimated number of punch cards
<u>J. FOETAL MORTALITY</u>		
J01	Late foetal deaths (total)	15,000
J02	" by age of mother (total, legitimate)	3,000
J03	" by age of mother and birth order	6,000
J04	" by sex	1,000
J05	" by legitimacy status	500
J06	Foetal deaths by gestational age	1,000
<u>K. PERINATAL MORTALITY</u>		
K01	Perinatal deaths	300
<u>L. LIFE TABLES</u>		
L01	Expectation of life at specified age for each sex	3,000
L02	Life table mortality rates at specified ages for each sex	2,000
L03	Survivors at specified ages for each sex	3,000
<u>M. MARRIAGE</u>		
M01	Marriages (total)	2,500
M02	" by age of bride and by age of groom	10,000
M03	" by age of bride classified by age of groom	2,000
M04	" by age and previous marital status of bride and groom	2,000
M05	" by previous marital status of bride classified by previous marital status of groom	3,000
<u>N. ANNULMENTS</u>		
N01	Annulments (total)	1,000

Table No.	Title	Estimated number of punch cards
<u>P. DIVORCE</u>		
P01	Divorces (total)	1,000
P02	" by age of wife classified by age of husband	500
P03	" by duration of marriage	1,000
P04	" by duration of marriage, classified by age of wife and age of husband	1,000
P05	" by number of dependent children	1,000
<u>Q. INTERNATIONAL MIGRATION</u>		
Q01	Long-term emigrants by age and sex	4,000
Q02	Long-term emigrants by country of intended residence	6,000
Q03	Long-term immigrants by age and sex	3,000
Q04	Long-term immigrants by country of last permanent residence	6,000
Q05	Major categories of departures and arrivals	3,000
Q06	Long-term intercontinental and continental migrants	400
<u>R. SEPARATION</u>		
	Separations (total)	500