



E/CEPAL/CDCC/41

15 March 1978

ECONOMIC COMMISSION FOR LATIN AMERICA
Office for the Caribbean

CARIBBEAN DEVELOPMENT AND CO-OPERATION COMMITTEE

Third Session
Belize City, Belize
12-18 April 1978

PROPOSAL FOR THE ESTABLISHMENT
OF A
STATISTICAL DATA
BANK



UNITED NATIONS

ECONOMIC COMMISSION FOR LATIN AMERICA Office for the Caribbean

9
3
10

1
12

PROPOSAL FOR THE ESTABLISHMENT OF A
STATISTICAL DATA BANK

BACKGROUND AND JUSTIFICATION

The Work Programme outlined for the Caribbean Development and Co-ordinating Committee's Secretariat at the Havana meeting of 1975 mandated the CDCC Secretariat to establish a number of measuring and monitoring activities that at base require the services of a statistical data bank and a documentation centre. The CDCC Secretariat therefore sees as one of its goals the development of a system of capture and storage of statistical data pertaining to the several CDCC Countries in the several areas of work as performed. The data system should ensure quick retrieval of data for use by planners and researchers, both government and private alike. The system should be able to generate quickly, tables in the form of time series, or other analytical tabular formats. The statistical data bank should complement the Documentation Centre, and statistical methodologies should find a place in the documentation records. The Documentation Centre was created to provide planners, policy-makers and researchers in the Caribbean with pertinent, accurate and timely information which would support the development and co-operation activities of CDCC member countries, and to promote and facilitate optimum utilization of information resources available within and outside the sub-region.

CONTENT OF THE DATA BASE

The statistical data base should contain time series published by or gathered from official data in each of the CDCC countries, and the other countries of interest.

The series should commence in 1970 or in the earliest year thereafter, for which continuous data are available. The following would represent a list of some of the sources that could be used to provide series for storage:- Business Statistics; Industry Statistics; Agriculture; Transport & Utilities; Mining; Manufacturing; Construction; Tourism; Distribution; General Statistics; External Trade; Prices; National Accounts and Balance of Payments; Household Statistics; Census Statistics; Health Statistics; Energy and Natural Resources.

The data base as suggested selectively includes the statistics at present collected and elaborated by the various sector specialists, which do not reach other officers who at times and at great cost to the office duplicate the effort of collection of data for use in studies. The inclusion of such data will bring about economies both in terms of time and expenditure.

ORGANIZATION AND STRUCTURE

The present lack of computer facilities has determined the nature of the system. The manual capture and storage of data will have to be adopted, but this must be done in such a manner as to facilitate an easy shift over to computerized systems when such become possible. The layout suggested in this paper attempts to meet the task of conversion to computer systems. It is proposed to record the data on cards 5" x 8" in size, which will in turn form part of a Kardex system. The storage should be alphabetical by country, with the entire data set for any given country being accessible immediately upon accessing the country. The cards should be numbered sequentially within country in order to provide a check on completeness of coverage for any country, and the numbering should be the same for each

country. In this way, accessing of comparable data for several countries could be made more direct and speedy by eliminating the need to go through sequentially a stack of cards in respect of a country. One possibility would be to number the countries according to the UN Country Codes as per Document M 49/Rev.1. This would provide an accepted code to be used as a country identifier. The series can also be assigned a numerical code. This will be of use to the computerization exercise that should follow. If the user wishes to obtain like data for all or some of the CDCC Countries, he can access the data by means of a subject code within country code sort.

The system should cater for two levels of confidentiality as it will records:

1. Annual data received from the annual survey, and at a later date, quarterly submissions of the various national agencies; and
2. data that have not been published either by the Governments or by CEPAL, but which at the same time are not regarded as being supremely confidential.

The latter category would represent the confidential/semi-confidential data which must be suitably codified and recognized to be classified data. If such data share the same storage area as the unclassified data, red lettering should be used in the manual system to warn the documentalist/internal user of the nature of the data. In the case of classified data received from a sector specialist, the documentalist/internal user is advised to consult the donor of the information to receive his written permission before divulging or publishing such data. In anticipation of the shift over to a mechanized system, the levels of

confidentiality can be indicated by means of a code to be assigned upon advice from a System Analyst, or preferably the systems analyst who will design the computer systems.

In Statistics, there is no such thing as a "final" figure. All figures are provisional and therefore subject to change in the light of new knowledge. In a time series, there will emerge different levels of acceptability of "firmness" of the data. The latter years' data might take the form of preliminary estimates, for which revisions are expected in the near future. The system of data capture should reflect this by recording the data to reflect their tentative nature. Whereas the use of a lead pencil is not considered to be highly desirable, in this case it might be acknowledged to be most conducive to a neat changeover to a firmer figure. Alternatively and more correctly, one can choose to delete the preliminary figure and insert the new.

It is inevitable that over a sufficiently long period of time the space on any given card will be exhausted, and there would exist the need to store and possibly re-access the summary data contained in the "bank". At this stage, even if the office does not possess its own computer or scheduled computer time, the data should be transferred to cards and/or tape and listed out according to any format that might be thought most useful. At around the same time, the cards can be microfilmed and disposed of. One problem in both operations mentioned would be the identification of classified data. The reproduction of the cards on computer tape and on microfilm is insensitive to colour coding, therefore another code, probably a numerical code, will have to be introduced along with the colour code.

Another problem caused by "dumping" the records on to microfilm is that it causes the coming into being of two records that must be consulted when a search has to be made - the microfilm record and the new cards with the more recent data. This split in the records impairs analysis, in that it increases processing time.

ADMINISTRATION

The Statistical Data Bank should be administered by the Statistician, in collaboration with the Librarian/Documentalist. These two officers are already on the staff, and will allocate a portion of their time to the project. The commitment of data to records should be done by Statistical clerks whose entire attention can be devoted to the exercise. Two clerks working for one year should be able to clear the seven-year backlog of data available.

RESERVATION

One serious qualification of the manual system as described is that in view of the rapid development of the technologies used in the processing, storage, retrieval and dissemination of information, and in view of the likely increase in the occurrences of multi-aspect queries over time, a manual search, both of statistical data as well as of document references, will become increasingly difficult. The logical option would favour a mechanized system. Discussion of this proposal may very well lead to a position on the part of Governments, to mechanize or not to mechanize the data base.

INPUTS FROM GOVERNMENTS

If the Governments opt for a mechanized system, there would be the need to finance the acquisition of a machine; additionally this acquisition would create an annabolism of wants. If, on the other hand, the option is for a manual system, the level of search and analysis will of necessity be adversely affected. In this case, the major contributions from the Governments would be the prompt supply to the CDCC Secretariat, via the national focal points, of the data needed for inclusion in the bank.

REQUIREMENTS OF A MECHANIZED SYSTEM

A mechanized system will necessitate the provision of both hardware and software, and systems and programming expertise. The size, cost and configuration of the hardware will be determined by the size of the files that will be created, and the type of data to be stored. A small machine will demand more systems and programming skills at the design stage. There will arise the need to retain the services of a systems analyst/programmer and a keypunch operator/machine operator at least. It is envisaged that the document references would be always "on the computer" to facilitate speedy referrals, without having to be loaded on and off with each request. The statistical data can be stored on tapes and accessed by whatever means may be advocated. At present, much in-depth analysis is not possible because of the difficulty of engaging in multivariate analysis. Although the onus and challenge will be on the sector specialists to utilize these tools, the availability of both the hardware and the appropriate software should result in a higher level of analysis of the CDCC countries by the CDCC Secretariat.

It is estimated that normal running time of the machine will be one shift of eight hours per day, with the possibility of work beyond the normal working hours when necessary.

RECOMMENDATION

Member governments are asked to consider the establishment of a Statistical Data Bank within the CDCC Secretariat. If this is agreed, members may wish to decide whether the system should be manual or mechanized. The tentative staff and equipment requirements are as follows:-

1. Manual System

Two (2) Statistical Clerks.

Two (2) filing cabinets.

Cards.

2. Mechanized System with a link to a Main Frame
Outside of the Office and not owned by
CDCC Secretariat

One (1) Keypunch operator/machine operator.

One (1) Statistical clerk.

One (1) Transaction System, one (1) tape unit
and one (1) Printer. (Present cost around
US\$60,000).

Tapes.

Computer time - An estimated US\$200.00 per hour; the
average number of hours to be utilized monthly = 6.

