



REDATAM

THE RELEVANCE OF THE REDATAM SYSTEM FOR THE 1990 CENSUSES

UNITED NATIONS
ECONOMIC COMMISSION FOR LATIN AMERICA AND THE CARIBBEAN
LATIN AMERICAN DEMOGRAPHIC CENTRE

Distr.
RESTRICTED
LC/DEM/R.48
Series A, No.174
September 1987
ENGLISH
ORIGINAL: SPANISH

THE RELEVANCE OF THE REDATAM SYSTEM FOR THE 1990 CENSUSES

ECONOMIC COMMISSION FOR LATIN AMERICA AND THE CARIBBEAN (ECLAC)

LATIN AMERICAN DEMOGRAPHIC CENTRE (CELADE)

Casilla 91 - Santiago de Chile

Tel: 2283206 Cable: UNATIONS

TABLE OF CONTENTS

SUMMARY	1
1. INTRODUCTION	2
2. THE REDATAM SYSTEM	3
3. EXPERIENCE WITH REDATAM	5
4. PLANNING THE 1990 CENSUSES FOR SMALL-AREA DATA UTILIZATION	6

THE RELEVANCE OF THE REDATAM SYSTEM FOR THE 1990 CENSUSES

SUMMARY

This document provides a non-technical description of the REDATAM 1/ software system for using ordinary IBM or compatible microcomputers to store the data of an entire population and housing census and to obtain any tabulations rapidly and at low-cost for any small geographical areas defined by the user such as cities, neighborhoods or city blocks or any combination of these. The system is particularly useful for regional planners and other non-programmers that require small area information.

The experience of using REDATAM with census data from Chile, Saint Lucia, Dominica, Costa Rica and Uruguay and with survey data from Guyana is briefly outlined along with a number of lessons that have been learned of relevance to other countries.

The implications of using the census with REDATAM for providing small area information are examined in terms of the planning and organization of the 1990 censuses. Particular emphasis is placed on good cartography, computerized if possible to permit easy updating and direct utilization with REDATAM, on the minimum geographic level at which data will be collected, on minimum comparable areas between the 1980 and 1990 censuses and on primary control to avoid errors that will be magnified when small area data are used.

In conclusion, it is noted that the REDATAM system can be used by a country to begin to provide timely results from its 1990 census to planners and other users long BEFORE the census tabulations are produced or the publications are ready for distribution and that, furthermore, the ease and speed of obtaining small area information with REDATAM can help reduce the number of geographically disaggregated tables that have to be published. It is also possible to use REDATAM to decentralize the information to regional and other sub-national offices.

1/ REDATAM = REtrieval of DATA for small Areas by Microcomputer.

THE RELEVANCE OF THE REDATAM SYSTEM FOR THE 1990 CENSUSES

1. INTRODUCTION

The objectives of this document are to describe experience gained in providing census information for small geographical areas and to indicate which stages of census processing can be improved to make the findings quickly available.

The word "census", usually brings to mind a tremendous and costly effort, and rightly so. In fact, the census operation is a long-term project (generally four to five years) which involves the integration of different fields, high costs, a huge personnel input, etc. Consequently, census information should be used exhaustively to improve the cost-benefit ratio.

Conceptually, the difference between census data and census information is characterized by two factors: a) information is extracted from the data after some kind of process, for example, the tabulation of results; and b) information must be available to the user (that is to say, even if census publications are available, if the user is unaware of them or if they do not satisfy his need, for him or her they remain just data and are not information).

The files created during processing are usually stored on magnetic tapes for purposes of convenience and because large amounts of storage are required. When tables have to be drawn up from these files, potential users encounter three very common problems in the statistical offices of Latin American and Caribbean countries: a) the processes are very costly because the entire census file has to be read; b) once the census is completed, the computers have their time scheduled for processing other services, or in the case of many Caribbean countries, are not available; and c) the programmers required are very busy with more urgent activities.

There are potential users who are systematically handicapped by all three limitations; these are municipal planning officials, investors, suppliers of social services and others who require information for small geographical areas. Almost always, the projects needing the information have limited budgets and are unable to cover the costs of reprocessing an entire census file, particularly since the work often has little or no importance in national terms. Consequently, the requests for the information are delayed or never answered.

THE RELEVANCE OF THE REDATAM SYSTEM FOR THE 1990 CENSUSES

The emphasis nowadays is on decentralizing and fractionalizing the planning process so that the number of potential census users increases as the universe being studied decreases. The result is that there is an increasing need for disaggregated census data that has been converted into information.

2. THE REDATAM SYSTEM

The REDATAM system (REtrieval of census DATA from small Areas by Microcomputer) was developed in CELADE 2/ to meet these unsatisfied user needs using an ordinary IBM microcomputer or compatible. The aim of REDATAM is "to organize and maintain geographically classified census files within the limited capacity of a microcomputer, so that tabulations or other statistics can be obtained for the smallest geographical unit, for example, a city or a block or any combination of these" 3/.

The system has been used so far to create the population and housing census databases of Chile (1982 census), Saint Lucia (1980), Dominica (1980), Costa Rica (1984) and Uruguay (sample of the 1985 census) and for a demographic survey of Guyana (carried out in 1986). The table below sets out the storage size and space required in each case.

Country	Dwellings (Thousands)	Persons (Thousands)	REDATAM space required (Megabytes)	Storage Method
Chile	4 000	12 000	300	Laser disk
Saint Lucia	30	125	3	Hard disk
Dominica	17	79	2.5	Hard disk
Costa Rica	500	2 500	60	Hard disk
Guyana (Survey)	8	42	1	Hard disk
Uruguay (Sample)	75	300	10	Hard disk

2/ With funds from the International Development Research Centre (IDRC) of Canada, and additional funding from the United Nations Fund for Population Activities (UNFPA) and the Canadian International Development Agency (CIDA).

3/ CELADE, REDATAM, User's Manual. CELADE, Santiago, May 1987.

THE RELEVANCE OF THE REDATAM SYSTEM FOR THE 1990 CENSUSES

A REDATAM data base needs approximately one fourth the space of the original files on magnetic tapes. In the case of Costa Rica, for example, the original files required 230 megabytes, filling six magnetic tapes of 2 400 feet each. This reduction (to 60 megabytes in the case of Costa Rica) and improvements in the speed of tabulation were obtained through of number of system design features, including:

- a) In ordinary sequential files, the geographical identification variables are repeated for each of the housing and population records. In REDATAM, this identification is required only once for each geographical level and a pointer system is used to access the actual data.
- b) In a sequential file, the information is stored according to the traditional method of one record per dwelling or person, which contains all the variables of the particular case (or observation). In REDATAM, an "inverted" file system is used, where one file is used for each variable, making it necessary to access only the variables of interest in any given tabulation.
- c) A data compression system is used, which only stores the necessary "bits". For example, the variable "sex" generally uses one byte (digit) to store the number 1 if the record is for a man, or 2 if it is for a woman. In the binary compression method used by REDATAM, the same codes 1 and 2 may be stored in 2 bits (a quarter of byte).

The REDATAM approach also yields results in terms of the time in which a user requirement or request is processed. For example, it is not necessary to read all the person variables if only a cross tabulation of age by sex is needed. All that is required is an input of the two variables and this renders execution much faster. In addition, by using pointers for the smaller geographical areas the entire file does not have to be read, that is, the computer can jump immediately to a group of city blocks, rather than passing through all the data as in a sequential file.

Obviously, there is no comparison between the present speed of a microcomputer and that of a mainframe computer; whereas in the former a process can take two hours, in the latter the same process might take only 5 minutes. However, because many users do not have easy access to a large computer and have to depend on programmers for using such computers, it is often more convenient to employ an easily available microcomputer than to wait days or

THE RELEVANCE OF THE REDATAM SYSTEM FOR THE 1990 CENSUSES

weeks for the large computer to become available. Moreover, in the vast majority of requests concerning small areas, the area of interest can be identified and the results obtained in a matter of minutes.

The REDATAM system has another feature of special importance to planners and many other users, namely, its hierarchical processing capability, whereby results involving both housing and population variables can be obtained, for example, tabulations of the number of primarily students by age and the size of their household, who live under given housing conditions.

3. EXPERIENCE WITH REDATAM

The Chilean and Saint Lucia databases were set up as a part of a pilot project and served to provide background information for developing the system. Both databases are being widely exploited by their respective users, who have made available their experiences in processing the data.

In the case of Saint Lucia, a country with 125 thousand inhabitants in 1980, the REDATAM system is used to process ALL requests for tables, since the statistical office does not have access to a large computer --the 1980 census was originally processed in Barbados--. One of the REDATAM requests, was made by the telephone company, when it wished to make a survey of a neighborhood of Castries, where it planned to expand telephone lines. Of interest for the 1990 census is that because the census had been taken at only three geographical levels (the province, the city and the enumeration district), the lowest level was not small enough to define exactly the area of interest to the company. Blocks or block faces would have been more useful for this purpose.

In Chile, a country with 12 million inhabitants, REDATAM is being utilized both by the National Statistical Institute (INE) and by CELADE and is almost always employed for obtaining information for small areas. A typical example is the study being done on the indigenous population in the comuna of Temuco, where the areas to be studied were first identified geographically. These areas were then defined systematically in the REDATAM system, and together they form the universe for the tabulations requested.

THE RELEVANCE OF THE REDATAM SYSTEM FOR THE 1990 CENSUSES

In Guyana, the system was used for data from a retrospective demographic survey (GUYREDEM, Retrospective Demographic Survey) carried out in 1986, and although the volume of data was not very large (42,000 cases), it served to prove that the system can be used with weighted files to take into account expansion factors. The great advantage of REDATAM in this application is the speed of processing and the small space occupied by the database when compared with the SPSS/PC (Statistical Package for the Social Sciences, Microcomputer Version). On the other hand, SPSS has much greater flexibility than the current REDATAM version since it can output tables that are ready for publication and it has more sophisticated statistical processes (however, when these features are required, REDATAM has an interface that permits it to output data from the REDATAM database directly to SPSS-PC).

The Guyanese case is also interesting because it presented identification problems in the questionnaires, since some were coded as belonging to enumeration districts that were not in the sample and, therefore, did not appear in the geographical files that are associated with REDATAM. These discrepancies between the identification of the questionnaires and their geographical location can result in serious errors when working with small areas.

4. PLANNING THE 1990 CENSUSES FOR SMALL-AREA DATA UTILIZATION

The concepts and the experience acquired with REDATAM should be incorporated into the decisions at the planning and organizational stages of the 1990 censuses, since these will affect the subsequent use of the census data for small areas.

In the first place, good cartography is absolutely necessary in order to identify the areas of interest correctly at the lower geographical levels. Hence, a good cartographic database, which can be updated during the periods between censuses, is essential for the exploitation of data for small areas. One way would be to maintain a computerized system for storing the maps, which would ensure full census coverage and automatic production of maps more quickly and accurately and permit in the future the combination of census and

THE RELEVANCE OF THE REDATAM SYSTEM FOR THE 1990 CENSUSES

cartographic data for presenting results in a graphic form 4/. The census done in Colombia in 1985 employed a computerized cartographic system called Micro-Map, that was developed internally in the national Administrative Department of Statistics (DANE) for the Apple IIe microcomputers 5/.

Prior definition of the minimum geographical identification level is very important because this will establish the limits of the census for obtaining small area information. For example, if the identification reaches only the city block level, it will not be possible to obtain data at the level of the city block face. The definition of the minimum geographical level has to be determined during the early planning stages because it will affect the cartography, the questionnaire design, the taking of the census, etc.

The primary control of the census questionnaires also plays an important role in the results obtained at the level of small areas. This process, which is executed when the data is input, is necessary in order to check inter alia the accuracy of the geographical location variables in questionnaires so as to ensure that these questionnaires are correctly entered for the smallest codified geographical levels (sector, city block, etc.).

The advantages of good cartography and proper boundary control are not fundamental for the country as a whole nor very important for large areas such as regions, inasmuch as the total populations are normally not significantly affected. However, any error in identifying a small area, such as a city block, could destroy users' hopes of obtaining useful information at this geographical level.

Another issue to consider involves historical comparability between the 1990 census and that of 1980. Comparability should be maintained, not only with respect to the basic questions in the census schedule but also with respect to the geographical subdivisions of the country. The concept of "minimum comparable areas" must be used rigorously, in other words, the minimum geographical identification with which the census results can be compared

4/ Silva, A., El Procesamiento de los Censos de Población en América Latina y el Caribe para la Década de 1990: Un Vistazo al Futuro. CELADE, Santiago, August 1986.

5/ Ferro Calvo, M. El Proyecto de la Cartografía Censal, Workshop to analyze and evaluate censuses, pp. 369-400, Buenos Aires, May 1985.

THE RELEVANCE OF THE REDATAM SYSTEM FOR THE 1990 CENSUSES

between two censuses should in fact, be the smallest possible, in order to enable disaggregated comparative studies between the successive censuses to be made.

Finally, it should be also mentioned that once the data are edited, the REDATAM system can actually begin providing public services BEFORE the census publications are produced or ready for distribution. Furthermore, the speed and ease with which information is obtained through REDATAM, at any geographical level, can have a positive impact on the number of tables to be published by limiting them to the more general results. At the same time there is a real possibility of disseminating census information at the regional level (decentralization of the information), by generating REDATAM sub-bases with data on the region in question and installed on microcomputer equipment located in the various regional offices.

(redcen.pub/celev2) (28.10.87)



