Distr.
RESTRICTED

LC/DEM/R.216
Series B, N°100
8 June 1994

ORIGINAL: ENGLISH

C E L A D E
Latin American Demographic Centre

POPULATION DATA FOR LOCAL DEVELOPMENT

This document was prepared by the Information and Technology Area of CELADE for publication in the September 1994 issue of the *FID News Bulletin* of the International Federation of Documentation special issue on "The POPIN Network and the relevance of Population Information for Development", which is part of the POPIN activities for the International Conference on Population and Development.

The text was not subject to editorial revision.

TABLE OF CONTENTS

- *The health authorities in a Peruvian municipality want to know whether poor single-parent women with young children have adequate access to existing maternal/child health clinic services, and want guidance on alternative sites for a possible new clinic.*

- *An educational authority in Chile is running a deficit since the supply of existing educational resources does not seem to correspond to the school age population; they want some alternatives on ways to bring supply closer to demand while offering better education.*

- *A company plans to build a new tourist complex on a small Caribbean island -- the Government wants to examine the likely effects on the environment and on the population who will be affected.*

Planning and decision making in each of these and a myriad of similar local and regional level problems involve the use of data concerning the distribution of population and other variables over space. While these are problems that affect both developed and developing countries, the focus in this paper is on the latter, and more specifically, on the Latin American and Caribbean (LAC) countries.

An increasing number of the LAC countries are coming to see that their efforts to achieve sustainable social and economic development requires decentralization of their societies with effective local capabilities for participation, planning and decision making. This in turn is stimulating many in the countries to look for and adopt relevant technological and methodological advances in the field of information.

Numerical vs. bibliographical data

In this paper we are concerned with numerical (including geographical) data and databases. As many readers may be more familiar with bibliographic data, a short digression will describe some germane differences between the *retrieval* of bibliographic and the *tabulation* of numerical data.

Assume I am a planner in a town concerned with improving social infrastructure to reduce poverty. It would be useful to *search* a large database that might have hundreds of thousands of entries to identify the 5 or 6 *specific* documents that provide pertinent information about the topic. However, if I want to know the distribution of poverty in the town, I shall probably use numerical data from the census or a special field study. The census has a record on the characteristics of each of the millions of persons living in a country, but I have no interest to identify specific individuals in the census, and, in fact, no Government will permit individuals to be identified in its census data. Rather, I want to *tabulate* the data on individuals in the town to *aggregate* them so that I can see, for example, the percentage of persons who are poor on each block in the town. I might use a census variable such as whether a person lives in a house without running water as an indicator of poverty.

To do this, I would use appropriate tabulation software with the census *microdata* (variables describing the characteristics and housing of each individual enumerated in the census) to divide the population living on each block into two categories: those who live in houses without running water and those that have water. The software will count the number of persons in each category on each block. Note that although the software must look at each "individual", what I need and will get is information about *categories* of individuals. The resulting table might look like:

Poverty in the Town of Miranda

Block ID	No.persons in deficient houses	No.persons in non-deficient houses
Block 001	78	15
Block 002	43	52
...	...	...
Block 284	97	39

With this information, I might then map the percentage of persons living in deficient houses on each block to see whether there are groups of contiguous city blocks --pockets of poverty-- with high proportions of people living in poverty.

With these clarifications, we now return to the problem of data for local development.

The development strategy and information access

The member Governments of the United Nations Commission for Latin America and the Caribbean (ECLAC) have agreed at regional fora that development in the 1990's and beyond should be based on the *transformation of production with social equity*. This approach includes such goals as:

a) Increased competitiveness in global markets through improved productivity via the incorporation of technological progress throughout the economy.
c) Continuous improvement in the quality of human resources.
d) Strengthening the political commitment of all sectors through fortifying participatory democracy and increasing equality.
d) Environmentally sustainable development.

Among the basic conditions for the progressive achievement of these broad policy objectives in the Latin American and Caribbean societies is the *spatial decentralization of power and resources*. To this end, various countries, among others, Bolivia, Brazil, Chile, Colombia, Venezuela and Trinidad and Tobago, are moving toward decentralization.

Effective decentralization within a country requires that *the public and private sectors at the regional and local levels have access to the data and the information tools needed for local planning and decision making.*

The census as the major source of small-area population microdata

Planning of social, economic and physical infrastructure at the local level normally requires geographically-referenced information, often for areas as small as city blocks. In most countries the only source of basic social and economic microdata for every area is the population and housing census. The existence of data for small-areas is perhaps the major comparative advantage of a census over a survey. For most other purposes, where high geographic "resolution" is not necessary, probabilistic sample surveys of a few thousand cases are cheaper, faster and more profound than a census with many millions of cases.

Yet until recently, census data was very much under-utilized in Latin America and the Caribbean, and, worst of all, given the very high cost of mobilizing a nation for a census, the censuses were least used where they had the most to offer -- for small-area data. There are various explanations, but here we are concerned with problems of technology.

The National Statistical Office (NSO) of a country, after carrying out the enumeration, normally makes the census results available in volumes of published tabulations produced from the microdata on reels of magnetic tape kept at the NSO. It is impossibly expensive and wasteful to try to anticipate *a priori* tabulations at the local level for the entire country. Imagine the amount of paper to provide the age distribution of mothers with small children living without a spouse for every city block of an entire country, just in case it is needed somewhere. And if a potential user then says that only illiterate women are of interest, all the existing paper is useless. It would have been much easier to make special tabulations, on request. But until the late 1980's, reprocessing many reels of magnetic tape to find the specific town of interest and then tabulating the information was difficult and time consuming.

Around 1988, the "reprocessing" of very large census files for these purposes became technically feasible on a microcomputer using software for the **RE**trieval of **DA**Ta for small **A**reas by **M**icrocomputer (Redatam), written by the United Nations Latin American Demographic Centre (CELADE). By compressing all the *microdata* of an entire census to 25% of its original volume without loss of information, and by organizing the data suitably (hierarchical transposed files), it became possible to obtain complex user-defined tabulations from very large census files of millions of cases for user-defined small areas in minutes instead of hours or days and at low cost in a user-friendly environment.[1]

By the end of the 1980's, over 30 LAC countries had Redatam census or other databases. Via training, technical assistance and networking, the NSO were stimulated to create 1980 census

[1] Redatam was created, and continues to be developed, with funds from the International Development Research Centre (IDRC) of Ottawa, Canada, with support from the Canadian International Development Agengy (CIDA), the United Nations Population Fund (UNFPA) and the UN Regular Budget.

Redatam databases so that they and potential users could become familiar with the technology for the then upcoming 1990 censuses.

The emergence of new information needs and associated technology

Geography

The utilization of data for small-area analysis inevitably requires maps of the areas involved. The municipal government cited previously, may want to know, for example, not just the average poverty level in its jurisdiction, but, if something is to be done, where the poverty areas are located within the territory. Thus, when the census comes to be seen as a major source of small-area population data, the census cartography becomes a necessary part of the data.

Furthermore, digitizing the census and other cartography so that it can be managed, updated and manipulated by software, now takes on growing importance. The usefulness of the small-area information, such as provided by Redatam, is much increased through *spatial display and analysis*; alternative scenarios can be *visualized* and compared to facilitate decision making. In the last few years, these operations have become possible on *geographic information systems (GIS)* working on microcomputers. Digitized databases for GIS are complex to create and relatively few exist at the municipal level, but this is clearly the direction of the near future in many LAC countries.

Multidisciplinary databases

As small-area population data have become available to municipalities and sectoral users, the number and range of users have been increasing. Most are secondary users of population data, that is, they work in other fields and require *multidisciplinary databases* which incorporate microdata from one or more censuses along with data from other sources in their own fields.

As part of its strategy of providing key technology for supporting decentralization, in 1992 CELADE released a new version, Redatam-Plus (R+), which generalized the previous system and added two major new features: the management and exploitation of multidisciplinary databases and an interface to allow R+ results to be transferred to a GIS for display and analysis, opening up the possibility of R+GIS applications. As far as known, Redatam-Plus remains to date the only microcomputer-based software that is optimized for obtaining complex small-area information from very large hierarchically-organized files of microdata.

By the end of 1993, R+, which is available in Spanish and English versions, had been distributed on request for a nominal price to nearly four hundred users throughout the world. Various countries, such as Bolivia, Chile, Colombia, Ecuador, Paraguay and Venezuela are actively considering or are engaged in making the census available to regional and/or municipal authorities as R+ databases (the software can easily download a R+ sub-database for any area from an entire census database). In Asia, Vietnam, with R+ translated into Vietnamese, and Malaysia are contemplating the use of R+ databases to help disseminate their censuses.

Solutions not software: Hiding R+ and GIS inside Windows applications

As the number and range of small-area population data users increases, new opportunities and

challenges are opening up within the framework of the development strategy of the transformation of production with social equity. Most of these secondary population data users need population data in certain aspects of their work, but are primarily concerned with other fields of interest like education, health and urban planning. They normally are not interested or prepared to use generalized software, but rather want "solutions" to their own problems.

Such potential users should have available R+GIS applications tailored to their own fields. An educational planner might want to *see* the distribution of students in an age category on a city map to consider alternative distributions of teaching resources in the existing schools, taking into account factors like nearness to pockets of poverty or student travel time. As most such users will have limited computer interest or skills, we do not want to burden them with learning R+ and even less a GIS; rather these should be hidden within a problem-oriented R+GIS application concerned with educational data and decisions. The more computer literate and interested could utilize advanced user options that permit the direct use of R+ and the GIS.

To these ends, CELADE and the University of Waterloo (UW) of Canada, which has expertise in GIS applications, have begun a new project, financed again by IDRC, to develop four R+GIS applications in the development fields of 1) Health and family planning; 2) Education; 3) Urban Planning; and 4) Tourism development impact on environment and population. Each application is being designed and will be tested to meet the needs of counterpart institutions in selected countries (Chile, Costa Rica and Saint Lucia), while being sufficiently general to be transportable elsewhere.

To facilitate the construction of powerful user-friendly R+GIS applications able to "hide" the R+ and the GIS systems working behind the scenes, CELADE is writing a new version of Redatam-Plus for Microsoft Windows 3.x. Tentatively known as *winR+*, it will have a Graphical User Interface (GUI) and take advantage of most Windows features.

In addition, there will be a *winR+GIS application developers workbench* that will allow programmers to write other applications according to the needs of their own end-user clients.

Institutionalization of Technology and Data Transfer

There is at least one more element to be considered in the dissemination of technology and data to help foster local initiatives. Simply providing data and software and letting actual and potential users "sink or swim" will not ensure effective widespread utilization. It is necessary to provide technical support to end-users, just as done by commercial software producers. However, neither CELADE nor UW have the resources to provide such support to a single country (e.g., Chile has around 350 municipalities) and much less to all the countries of the Region).

While there are various models for institutionalizing support to data and technology users, and the situation will vary by country, the most appropriate initially would seem to be that of national "self-reliance", that is, having an appropriate decentralized governmental agency within a country provide the support. This might be the National Statistical Office (NSO) since it is normally the owner of the census data and the most knowledgeable on their use. In this model,

the central NSO would train technical staff in its regional offices, which in turn would train municipal and other persons, as well as providing local technical support for the software and data.

The development strategy based on the transformation of production with social equity endeavours to create systematic changes throughout the society. For this reason, the utilization of multidisciplinary databases with R+GIS applications in well-endowed municipalities is at best only a start. Hence, institutionalizing the use of census and other population data with suitable technology at the local level is now a major challenge in the field of population information. Social as well as technical inventiveness at the national and international levels will be required.

Contact persons:
Arthur Conning, Chief, Information and Technology Area
Serge Poulard, Head, Data Processing Unit
CELADE
Casilla 91, Santiago, CHILE
Fax: (562) 208-0252

Prof. G.Brent Hall
Associate Professor
Faculty of Environmental Studies
University of Waterloo
CANADA
Fax: (519) 725-2827

r3PopFI5.pub CEamc103 (23.3.94 <17:55> ¦)