

Distr.
RESTRINGIDA
LC/DEM/R.56
Serie A, No.183
30 de junio de 1988
ORIGINAL: ESPAÑOL

**REDATAM: UNA HERRAMIENTA
PARA OBTENER INFORMACION DESAGREGADA GEOGRAFICAMENTE
PARA LA PLANIFICACION**

Documento presentado en el Seminario Latinoamericano de Recolección
y Procesamiento de Datos Demográficos
Santiago de Chile, 23 al 27 de mayo de 1988

COMISION ECONOMICA PARA AMERICA LATINA Y EL CARIBE (CEPAL)
CENTRO LATINOAMERICANO DE DEMOGRAFIA (CELADE)
Casilla 91 - Santiago de Chile
Telef: 2283206 Cable: UNATIONS

TABLA DE CONTENIDO

I. INTRODUCCION	1
II. EL SISTEMA REDATAM	3
III. RESUMEN DE LAS FACILIDADES DEL SISTEMA REDATAM	4
IV. LIMITACIONES DEL SISTEMA	7
V. EXPERIENCIAS EN EL USO DEL REDATAM	8
VI. REDATAM EN EL FUTURO	11
VII. CONCLUSIONES	13

I. INTRODUCCION.

1. El objetivo de este documento es presentar la experiencia adquirida por el CELADE en la utilización de la información censal a través de un sistema computacional que permite explotar los datos especialmente a nivel de áreas geográficas pequeñas, tales como ciudades, barrios o manzanas, o cualquier agrupación de éstas, a un costo reducido y directamente por el usuario por medio del uso de los microcomputadores.

2. En junio de 1983, un experto del Centro Latinoamericano de Demografía (CELADE), visitó seis países de América Latina y el Caribe -Brasil, Costa Rica, Chile, Perú y Trinidad y Tabago- con el propósito de examinar la situación de los usuarios de los datos estadísticos. 1/

3. Como resultado de la misión, se comprobó la existencia de serias dificultades en términos de la disponibilidad de la información sobre población que se describen a continuación en los siguientes párrafos: 2/

a) El desarrollo de nuevos proyectos o el mejoramiento de los servicios sociales tiene una dimensión espacial que se refleja en la localización de los mismos. Dentro de los muchos factores que se deben considerar al efectuar la planificación de tales proyectos, están las características y la distribución geográfica de la población que será beneficiada y/o afectada por los mismos.

b) Los datos de población de uso más frecuente en el país provienen de los censos de población y vivienda, pero, en general, las áreas geográficas del proyecto difícilmente corresponden a los límites político-administrativos; ello hace que el usuario se vea en la necesidad de "construir" su área de interés a través de una composición de las subdivisiones censales, tales como los distritos de enumeración.

1/ Conning, Arthur M., Report to IDRC of the REDATAM Pre-Project Mission, CELADE, Santiago, Chile, octubre 1983.

2/ Conning, Arthur M., Geographically Disaggregated Census Data for Planning in Developing Countries, International Union for the Scientific Study of Population Conference, Florencia, Italia, 1985.

c) Es prácticamente imposible que las oficinas de estadística publiquen (o almacenen los informes del computador) la totalidad de la información sobre las áreas más pequeñas del país. Suponiendo que un usuario requiera un conjunto de cuadros de un área específica que no corresponda a ninguna de las tabulaciones existentes (o a una combinación de ellas), la única solución es reprocesar los datos censales de acuerdo con las necesidades específicas.

d) La mayoría de las oficinas de estadística de la región no están equipadas para reprocesar los datos censales en forma rápida y a un bajo costo; en parte, porque los datos son voluminosos y están organizados convencionalmente (en cintas magnéticas). Además, una vez terminado el censo tanto computadores como programadores tienen su tiempo dedicado a otros servicios de más alta prioridad. Las solicitudes de procesamientos especiales para áreas pequeñas deben, generalmente, esperar semanas o meses y a veces tienen un costo total muy elevado.

4. Existe un tipo de usuario potencial que resulta sistemáticamente perjudicado por las cuatro características mencionadas: los que requieren información sobre áreas geográficas pequeñas, tales como las oficinas municipales de planificación, los inversionistas, los proveedores de servicios, etc. Casi siempre se trata de proyectos con escasos recursos, sin posibilidad de hacer frente a los costos de procesamiento del archivo censal, con poca o ninguna relevancia en términos nacionales y sus demandas suelen quedar insatisfechas.

5. La tendencia del mercado de los usuarios del censo es la de multiplicar su cantidad al disminuir el universo en estudio. El énfasis actual está en la descentralización y en la explosión, o pulverización de los actores a niveles menores del proceso de planificación.

II. EL SISTEMA REDATAM.

6. Con el objeto de solucionar los problemas enumerados y ante la ausencia de opciones en el mercado se decidió construir un sistema especial ad hoc. ^{3/} Para una mayor flexibilidad de su uso, se optó por almacenar todos los microdatos censales (es decir, el valor directo de cada variable - sexo, edad, ocupación, etc. - para cada individuo o registro censal). La gran ventaja de esta estrategia es que es posible elaborar cualquier cuadro, independiente del nivel de detalle, para el área o grupo de áreas que se solicite. La posible desventaja sería el tamaño de esta base de datos, la que podría ser superada con una cuidadosa técnica de compresión en su almacenamiento.

7. Tomando en cuenta las limitaciones (tiempo y costo) de acceso a los grandes computadores y a sus programadores, un componente central de la solución planteada lo constituye el uso de los microcomputadores, de bajo costo y de uso corriente (IBM o compatibles), eliminando así la dependencia de los usuarios de los recursos computacionales centrales de las oficinas de estadística. Estos equipos se vuelven cada vez más baratos, están disponibles en la región y pueden ser usados directamente por los propios investigadores y por otras personas que no necesariamente sean programadores.

8. La filosofía existente detrás del REDATAM se refleja en el espacio necesario para almacenar una base de datos que es aproximadamente cuatro veces menor que el espacio inicial ocupado por los archivos en cinta magnética y en el tiempo de procesamiento de las solicitudes de los usuarios. Por ejemplo, si se requiere un tabulado de edad por sexo, no es necesario leer todas las variables. Sólo se necesita ingresar las variables involucradas, lo que aumenta considerablemente la velocidad de ejecución. Además, al utilizar los punteros para las áreas geográficas más pequeñas, no es necesario recorrer todo el archivo para procesar, por ejemplo, los datos de un conjunto de manzanas.

^{3/} Con fondos del Centro Internacional de Investigaciones para el Desarrollo (CIID) de Canadá, y el apoyo adicional del Fondo de las Naciones Unidas para Actividades en Materia de Población (FNUAP) y de la Agencia Canadiense para el Desarrollo Internacional (ACDI).

9. No se puede comparar la velocidad de los microcomputadores con la de los equipos más grandes, en los cuales un proceso puede durar dos horas, siendo que en los primeros quizás sólo tardaría cinco minutos. Sin embargo, la dificultad de acceso de los usuarios a los equipos grandes y la dependencia de los programadores, hace que sea preferible utilizar el microcomputador que está al alcance de la mano ante la otra alternativa de esperar días o semanas a que el equipo grande esté disponible. Por otra parte, en la gran mayoría de las solicitudes de datos sobre áreas pequeñas, no se necesita más que unos cuantos minutos para identificar el área de interés y obtener los resultados.

10. Una ventaja adicional del sistema REDATAM es su capacidad de procesamiento jerárquico, que permite obtener resultados cruzados entre variables de vivienda y población como, por ejemplo, el número de estudiantes por sexo y tamaño del hogar que vive bajo determinadas condiciones referidas al estado del techo y paredes de la vivienda.

11. La versión 1.00 del sistema REDATAM fue instalada en los dos países piloto (Chile y Santa Lucía durante el año de 1986). La versión 2.00 comenzó a distribuirse en abril de 1987 y actualmente está disponible la versión 3.00, recién terminada en marzo de 1988. Existen tres manuales de utilización del sistema: a) Manual del Usuario 4/; b) Manual de Generación de Bases de Datos 5/; y c) Manual de Facilidades de la nueva Versión 3.00 6/. Tanto el paquete como los manuales se encuentran disponibles en Español y en Inglés.

III. RESUMEN DE LAS FACILIDADES DEL SISTEMA REDATAM.

12. Selección geográfica

Definición del universo de casos para procesamiento, seleccionándose solamente las áreas geográficas pertinentes.

4/ CELADE, REDATAM, Manual del Usuario, (LC/DEM/G.50), Santiago, mayo de 1987.

5/ CELADE, REDATAM, Manual de Generación de Bases de Datos, (LC/DEM/G.53), Santiago, octubre de 1987.

6/ CELADE, REDATAM, Facilidades de la Nueva Versión 3.00, a ser publicado.

13. Agrupación de áreas geográficas

Posibilidad de componer conjuntos geográficos de interés a partir de áreas pequeñas aunque éstas no estén bajo la misma jerarquía geográfica, como en el caso de las áreas metropolitanas, o el estudio de municipios de dos departamentos distintos pero que utilizan una misma carretera (o bañados por un mismo río).

14. Base de datos autodocumentada

Las variables tienen un nombre corto (usado en la programación); un título (para despliegue en los resultados); definición de las categorías existentes (con códigos y nombres) y otros parámetros técnicos.

15. Utilización interactiva o en forma encadenada

Funciona en base a menús y con mucha ayuda en línea para el usuario. También detecta muchos de los errores de la digitación.

Se puede también pre-programar varios procesos largos y ejecutarlos de una forma encadenada, sin la intervención humana y en un horario adecuado (en la noche, por ejemplo).

16. Cálculo de variables derivadas

Utilización de las operaciones aritméticas e instrucciones de recodificación para la obtención de nuevas variables, como el grupo quinquenal de edad, o la calidad de la vivienda a partir de la estructura del techo, piso y pared. Estas pueden ser temporales (durante el proceso) o incorporadas a la base de datos.

17. Resultados estadísticos

Producción de tres tipos de estadísticas: frecuencias, tabulaciones cruzadas y promedios, estos dos últimos con hasta cuatro variables. Es posible obtener resultados porcentuales por filas, columnas o por totales, así como resultados para el total del área de interés y/o para niveles geográficos intermedios.

18. Procesamiento jerárquico

Existencia de dos niveles de variables (vivienda y población), con obtención de resultados para cada uno de los niveles, o para ambos a la

vez. Por ejemplo, el número de personas por sexo y grupos de edades que viven bajo determinadas condiciones de techo y pared.

19. Generación de sub-bases

Facilidad para crear pequeñas bases de datos REDATAM que incluyen solamente las áreas de interés de un determinado estudio. A partir de la base a nivel nacional, se pueden crear, por ejemplo, bases de datos por departamento y éstas ser entregadas a los centros regionales.

20. Utilización de factores de ponderación

Aplicación de pesos a los resultados estadísticos con el fin de obtener totales ponderados como en los censos por muestra o en las encuestas de hogares.

21. Producción de archivos para otros paquetes

En el supuesto de que las operaciones estadísticas básicas no sean suficientes, se pueden generar archivos con las variables de interés y con las observaciones correspondientes al área seleccionada, para ser procesados posteriormente por otro paquete como el SPSS o el SL-MICRO.

22. Almacenamiento de resultados

Se pueden guardar las salidas de los procesos estadísticos para: su posterior impresión, su incorporación en textos de documentos, así como para la producción de gráficos, usando otros paquetes tales como Wordperfect, Wordstar, Lotus, Mathplan, etc.

23. Creación de bases de datos en forma descentralizada

Posibilidad de crear bases de datos parciales (regiones, por ejemplo), y de conectarlas al final, facilitando el proceso en el caso de los archivos voluminosos.

24. Protección de la información

Implementación de los conceptos de secreto de la información y protección de los datos de la base, incluyendo "passwords" ("claves") para las funciones de administración y para el acceso a los distintos niveles

geográficos, con el fin de respetar la confidencialidad e impedir la identificación de viviendas y personas por usuarios no autorizados. Protección también para la creación de sub-bases de datos, con la eliminación, a elección, de los niveles más bajos de la jerarquía geográfica, posibilitando así la exportación de sub-bases de datos con la debida protección de la confidencialidad.

25. Variables decimales

Implementación del almacenamiento y cálculo de variables decimales.

IV. LIMITACIONES DEL SISTEMA.

26. Equipos IBM

El sistema fue diseñado para ser operado en equipos IBM/PC/XT/AT o compatibles. Ya se probó su ejecución en equipos EPSON, TOSHIBA, COMPAQ y GREAT WALL (computador chino).

27. Solo dos niveles jerárquicos

Si el censo tiene tres niveles (vivienda, hogar y personas), se genera hogar y personas como primer y segundo nivel respectivamente y, para la obtención de resultados a nivel de vivienda, se seleccionan sólo los primeros hogares. Para otro tipo de archivos con más de dos niveles no hay solución.

28. Tiempo de procesamiento para selecciones geográficas mayores

No se puede comparar la velocidad de los microcomputadores con la de los equipos grandes ("mainframe"). Para un país con más de un millón de habitantes y en que el universo de interés es todo el país, es preferible utilizar el equipo central.

29. La creación de una base de datos no siempre es una tarea sencilla

Para censos y encuestas relativamente pequeñas en los que es factible el almacenamiento inicial de los datos en forma convencional en un microcomputador, la creación de la base REDATAM es relativamente sencilla. Sin embargo, para los archivos más grandes, el proceso resulta complejo y

se requiere de un programador con conocimientos del equipo central y del microcomputador. En ambos casos, es imprescindible la existencia de una línea de comunicación entre los dos equipos.

30. Falta de integración con una base de información cartográfica

La selección geográfica tiene que hacerse en base a los códigos y nombres de las localidades (departamentos, comunas, distritos, etc), lo que exige una previa consulta de los mapas para definir el conjunto de códigos a ser utilizado.

V. EXPERIENCIAS EN EL USO DEL REDATAM.

31. El sistema REDATAM ha sido utilizado en varios países. La tabla 1, que se presenta a continuación, detalla los tamaños y espacios de almacenamiento requeridos en cada caso y a continuación se describen brevemente los detalles de cada uno de los países, 7/ y 8/, con algunos comentarios según el caso.

Tabla 1: Almacenamiento requerido para bases de datos REDATAM

País	Viviendas (miles)	Personas (miles)	Espacio (Megabytes)	Medio de almacenamiento
Chile	4.000	12.000	300	Disco Laser
Santa Lucía	30	125	3	Disco Duro
Costa Rica	500	2.500	60	Disco Duro
Guyana (encuesta)	8	42	1	Disco Duro
Uruguay (muestra)	75	450	10	Disco Duro
Dominica	17	74	2,6	Disco Duro
Colombia (básico)	5.800	27.800	70	Disco Duro
Colombia (ampliado)	600	2.800	60	Disco Duro

7/ Silva, Ari, REDATAM: Relevancia para los Censos de 1990, (LC/DEM/R.48), CELADE, Santiago, Chile, septiembre de 1987, 9 pp.

8/ Silva, Ari, REDATAM: Un Resumen, (LC/DEM/R.50), CELADE, Santiago, Chile, septiembre de 1987, 12 pp.

SANTA LUCIA

Base de datos del censo de 1980.

Obtención de resultados a nivel de país para pedidos de diferentes usuarios (Ministerio de Salud, Agricultura, etc.).

Estudio de la expansión telefónica para algunos barrios de Castries.

Estudio para la construcción de una represa.

Aplicación al proyecto de expansión de la red de alcantarillado.

URUGUAY

Base de datos de la muestra del 15 por ciento para resultados anticipados del censo de 1985.

Proyecto para el estudio de la pobreza crítica de todas las regiones del país, incluyendo características de la vivienda y de la población.

Tabulaciones de verificación para la publicación de resultados avanzados de la muestra.

COLOMBIA

Dos bases de datos, una para el censo completo de 1985 (formulario básico), y otra para el formulario ampliado del mismo censo.

33. En el caso de Santa Lucía, un país que no dispone de un computador grande y cuyo censo fue procesado en Barbados, el sistema REDATAM se utiliza para TODAS las solicitudes de tabulados.

34. En Guyana, el sistema se usó para cargar los datos de la Encuesta Demográfica Retrospectiva (GUYREDEM, Retrospective Demographic Survey) de 1986, que a pesar de no ser muy significativa en términos del volumen de la información, sirvió para poner a prueba el sistema con archivos ponderados (factor de expansión).

35. En el caso de Costa Rica, es interesante comentar que algunos usuarios externos a la Dirección General de Estadística y Censos (por ejemplo, la Compañía de Luz), se mostraron muy interesados en tener un subconjunto de variables de la base de datos del censo que pudiera ser procesado en sus propios microcomputadores. Este proyecto está en estudio en la Dirección de

Estadística. También, se generó una base de datos para la Encuesta de Hogares de 1986 y aunque los factores de expansión no permitían que se obtuvieran resultados de áreas muy pequeñas por tratarse de una encuesta por muestreo, fue posible procesar un cruce de variables para toda la encuesta en menos de cinco minutos. Otro archivo creado en REDATAM fue el del Registro Civil con los nacimientos y defunciones.

36. Como en el caso de Colombia el censo de 1985 se hizo por muestra, se generaron dos bases de datos, una para el universo del país y otra para el cuestionario ampliado, ambas con la misma restricción de que con el cuestionario ampliado no se pueden obtener resultados de áreas muy pequeñas porque los factores de expansión no son confiables a estos niveles. Un detalle interesante en el caso colombiano es que en verdad se generaron seis bases de datos regionales, de manera que se puedan usar en forma descentralizada en las oficinas regionales del Departamento Administrativo Nacional de Estadística (DANE).

VI. REDATAM EN EL FUTURO.

37. Los técnicos del Instituto Nacional de Estadística, Geografía e Informática (INEGI) de México están experimentando la generación de una base de datos REDATAM para el censo económico.

38. Existen convenios en marcha para la generación de bases de datos censales para varios países de la región como Bolivia, Cuba, Trinidad y Tabago y Venezuela. De éstos el más concreto es el de Trinidad y Tabago, donde se pretende generar la base de datos en este semestre.

39. Las autoridades brasileñas involucradas en el censo de 1990, están considerando la utilización de REDATAM para la obtención rápida de resultados del censo piloto a ser realizado en septiembre de 1990. Además, tienen también la intención de hacer una distribución masiva de los datos del censo de 1990 para uso público en forma de bases de datos REDATAM, los que serían enviados

junto con una copia del paquete, permitiendo a los usuarios trabajar directamente con los datos en sus propios microcomputadores.

40. El CELADE presentó al IDRC un proyecto para desarrollar en un período de dos años la segunda etapa del proyecto REDATAM, a ser conocido como REDATAM+. El nuevo sistema tiene como objetivo solucionar las necesidades de un "banco de datos para la planificación", capaz de almacenar y recuperar la información multidisciplinaria asociada con datos de población y vivienda, con facilidades de manipulación cartográfica y visualización espacial de la información, particularmente en los niveles regionales y locales. También el REDATAM+ está dirigido a garantizar que los censos de 1990 de América Latina y el Caribe sean capaces de proveer la información necesaria para estos propósitos. El proyecto fue aprobado y sus actividades iniciales comienzan en abril de 1988.

41. Estos objetivos reflejan una tendencia general en la región, cual es la de concentrarse en la planificación sectorial local y orientada espacialmente, con la visión de que los enormes gastos requeridos para la ejecución de un censo tengan un mejor retorno y una mayor utilización de su información que en el pasado.

42. Los grandes objetivos del REDATAM+ son los siguientes: a) Relacionar datos sobre población y otros temas con la información cartográfica a través de la creación de un Sistema de Información Geográfica (conocido como GIS en Inglés); b) Asociar un tipo de información "ecológica" multidisciplinaria sobre la descripción de áreas geográficas con datos de población y vivienda; c) Operar con la misma base de datos en una red de información local ("Local Area Network"); y d) Producir tabulaciones y resultados listos para impresión ("camera ready").

43. Lo que hará del REDATAM+ una herramienta distinta de las demás es que la información "ecológica" ^{9/} que describe las áreas geográficas pueden ser relacionadas, de acuerdo a las necesidades del usuario, con el microdato

^{9/} Para los propósitos aquí, "datos ecológicos" son las variables que describen un área geográfica, tales como el sueldo promedio de un municipio, o si una manzana cuenta con una clínica de salud. Por otro lado, datos "individuales" son las variables que describen personas, viviendas u otras unidades que son de interés solamente cuando agregadas estadísticamente.

"individual" de población y vivienda los que a su vez pueden ser procesados con las especificaciones del usuario (de hecho, los datos ecológicos, podrían haber sido generados a partir de los datos individuales, a través de agregaciones en los niveles geográficos superiores).

VII. CONCLUSIONES.

44. En resumen, el sistema REDATAM tiene como propósito "organizar y mantener los archivos censales, geográficamente ordenados, dentro de la capacidad limitada de un microcomputador, de tal modo que se puedan generar tabulaciones u otras estadísticas para la unidad geográfica más pequeña, como por ejemplo barrio o manzana o para cualquier agrupación de éstas". Así, los archivos censales que antes sólo podían utilizarse en grandes computadores, actualmente se pueden almacenar y procesar en microcomputadores al alcance del usuario.

45. De esta manera, no existe la intención de competitividad con otros paquetes existentes en el mercado para procesamiento de las diversas etapas de censos y encuestas. El REDATAM NO es un paquete de "procesamiento" en el sentido de "producir" o "transformar" la información de los cuestionarios. NO es para la entrada de datos, consistencia y producción de tabulados "camera ready". Su ambiente, tal como se ha descrito en el párrafo anterior, puede ser traducido más como una herramienta de almacenamiento de grandes volúmenes de información, de fácil y rápido acceso directo por el usuario final de la información, y a un bajo costo.

46. En el caso de los censos de la década del 90, aún si no se utiliza el sistema REDATAM frente a la eventualidad de que surjan herramientas más modernas, los conceptos y la experiencia adquirida en su utilización deberían incorporarse a las decisiones sobre planificación y organización, las que a la larga afectan la posterior utilización de los datos para áreas pequeñas.

47. A pesar de estar dirigido a la obtención de resultados para áreas geográficas pequeñas, el REDATAM puede ser de gran valor en el análisis de las encuestas, siempre y cuando sean respetados los límites impuestos por la metodología de la muestra. Es decir, las áreas seleccionadas para el proceso

deben contener un número mínimo de casos para que el resultado sea significativo.

48. Finalmente, cabe destacar que el sistema REDATAM, una vez creada la base de datos, puede prestar servicios ANTES que las publicaciones del censo estén listas. Por otra parte, la facilidad y rapidez en la obtención de la información para cualquier nivel geográfico, puede tener un impacto positivo en la cantidad de cuadros a ser publicados, limitándolos a los resultados más generales. Al mismo tiempo, existe la posibilidad real de difundir la información censal a nivel regional (descentralización de la información), mediante la generación de sub-bases REDATAM, con datos sobre la región de interés e instaladas en equipos microcomputarizados, localizados en forma descentralizada.

ref:\asat\red\texto\iusp.sem

