



REDATAM

REDATAM: RELEVANCIA PARA LOS CENSOS DE 1990

NACIONES UNIDAS
COMISION ECONOMICA PARA AMERICA LATINA Y EL CARIBE
CENTRO LATINOAMERICANO DE DEMOGRAFIA

Distr.
RESTRINGIDA
LC/DEM/R.48
Serie A, No.174
Septiembre de 1987
ORIGINAL: ESPAÑOL

REDATAM: RELEVANCIA PARA LOS CENSOS DE 1990

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

TABLA DE CONTENIDO

RESUMEN	1
I. INTRODUCCION	2
II. EL PROYECTO REDATAM	3
III. EVALUACION	6
IV. CONCLUSIONES	7

REDATAM: RELEVANCIA PARA LOS CENSOS DE 1990

RESUMEN

El propósito de este documento es describir un sistema que permite explotar, a un costo reducido, los datos censales a nivel de áreas geográficas pequeñas.

El proyecto REDATAM¹ constituye una herramienta para la utilización de microcomputadores en la producción de resultados derivados de la boleta censal, principalmente en lo que se refiere a las áreas geográficas menores, tales como ciudades, barrios o manzanas, o cualquier agrupación de éstas.

Se identifican los problemas generales en la obtención de la información censal y el perfil de los usuarios de datos de áreas pequeñas, tales como planificadores municipales.

El análisis de la experiencia en la utilización del REDATAM en los diversos países, tiene por objeto prestar ayuda a la planificación de los censos futuros, al reflejar algunos de los problemas enfrentados en la profundización de los niveles de explotación censal.

En la conclusión, se enfatiza la necesidad de contar con un buen sistema cartográfico, en lo posible computarizado y un adecuado control primario de las boletas censales.

¹ REDATAM = REcuperación de DATos para Areas Pequeñas por Microcomputador.

REDATAM: RELEVANCIA PARA LOS CENSOS DE 1990

I. INTRODUCCION.

El objetivo principal de este documento es presentar la experiencia adquirida en la utilización de la información censal para áreas geográficas pequeñas y establecer cuáles son las etapas del plan de procesamiento de los datos censales que influyen tanto en la calidad como en la rápida disponibilidad de los resultados.

Cuando se menciona la palabra CENSO, generalmente se piensa, con razón, en un esfuerzo grande y costoso. En efecto, se trata de un proyecto de largo plazo (cuatro a cinco años), que involucra la integración de áreas diversas, costos elevados, gran cantidad de personal, etc. Por esta razón, la información censal debería usarse en forma exhaustiva, a fin de mejorar la relación costo-beneficio del censo.

Conceptualmente, la diferencia entre datos censales e información censal se caracteriza por dos factores: a) la información se deriva de los datos una vez que éstos han sido sometidos a algún tipo de procesamiento (por ejemplo, la tabulación de resultados); y b) la información la constituyen los datos procesados, al alcance del usuario (es decir, aunque se disponga de publicaciones censales, si el usuario no las conoce o si no satisfacen sus requerimientos, continúan siendo datos y no información).

Los archivos generados como resultado del procesamiento, o la base de datos censales, generalmente se almacenan en varias cintas magnéticas, por razones de conveniencia y limitación de espacio en los medios de grabación. Cuando se trata de obtener algunas tabulaciones de estos archivos, los potenciales usuarios se enfrentan a tres problemas comunes a las oficinas de estadística de los países de América Latina: a) los costos son muy altos porque involucran la lectura de todo el archivo censal; b) una vez concluido el censo, los computadores son destinados a otros servicios; y c) los programadores están muy comprometidos en tareas de más alta prioridad.

REDATAM: RELEVANCIA PARA LOS CENSOS DE 1990

Existe un tipo de usuario potencial que resulta sistemáticamente perjudicado por las tres limitaciones mencionadas: los investigadores 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 recursos escasos, imposibilitados 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.

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

II. EL PROYECTO REDATAM

El proyecto REDATAM (REcuperación de DATos censales de Areas pequeñas por Microcomputador), fue desarrollado por CELADE ² frente a las necesidades no satisfechas de los usuarios. Su propósito es "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". ³

Hasta el momento el sistema ha sido utilizado para generar las bases de datos de los censos de Chile de 1982, Santa Lucía de 1980, Costa Rica de 1984, la muestra del censo del Uruguay de 1985 y una pequeña encuesta demográfica de

² 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 (UNFPA) y de la Agencia Canadiense para el Desarrollo Internacional (ACDI).

³ CELADE, REDATAM, Manual del Usuario, (LC/DEM/G.50), Santiago, Mayo de 1987.

REDATAM: RELEVANCIA PARA LOS CENSOS DE 1990

Guyana de 1986. El cuadro a continuación detalla los tamaños y espacios de almacenamiento requeridos en cada caso.

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	300	10	Disco Duro

El espacio necesario para una base de datos REDATAM es aproximadamente cuatro veces menor que el espacio inicial ocupado por los archivos en cinta magnética. En el caso de Costa Rica, por ejemplo, estos archivos ocupaban seis cintas magnéticas de 2.400 pies cada una, totalizando cerca de 230 Megabytes. Esta reducción considerable es producto de algunos rasgos internos del sistema, entre los cuales los más importantes son:

1. en los archivos secuenciales comunes, las variables de identificación geográfica se repiten para todos los registros de vivienda y población. En REDATAM, esta identificación se guarda sólo una vez para cada nivel geográfico, utilizándose un sistema de punteros para tener acceso a los datos propiamente tales;
2. en los archivos secuenciales, la información es almacenada según el método tradicional de un registro para cada vivienda o persona, que contiene todas las variables del caso (u observación). En REDATAM, se usa el sistema de archivos "invertidos", con un archivo para cada variable, a la vez que un sistema de compresión de datos, almacenándose sólo los "bits" necesarios. Por ejemplo, la variable SEXO usa generalmente un byte

REDATAM: RELEVANCIA PARA LOS CENSOS DE 1990

(dígito) para guardar los números 1 si es hombre o 2 si es mujer. En el método binario de compresión, los mismos códigos 1 y 2 pueden ser almacenados en dos bits (una cuarta parte del byte).

La filosofía tras el REDATAM también se refleja 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 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.

No se puede comparar la velocidad de los microcomputadores con la de los equipos más grandes, en los cuales un proceso que en aquéllos puede durar dos horas, quizás sólo tardaría cinco minutos. Sin embargo, la inaccesibilidad del común 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 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 necesitan más que unos cuantos minutos para identificar el área de interés y obtener los resultados.

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.

Así, los archivos censales que antes sólo podían utilizarse en grandes computadores, actualmente pueden almacenarse en microcomputadores al alcance del usuario común.

REDATAM: RELEVANCIA PARA LOS CENSOS DE 1990

III. EVALUACION.

Las bases de datos de Chile y de Santa Lucía se crearon como parte del proyecto piloto y sirvieron como antecedentes del desarrollo total del sistema. Ambas siguen siendo ampliamente utilizadas por los respectivos usuarios, quienes brindaron toda su experiencia en el procesamiento de las mismas.

En el caso de Santa Lucía, un país con 125 mil habitantes en 1980, 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. Una de estas solicitudes, por ejemplo, provino de la compañía de teléfonos, con el objeto de estudiar de uno de los barrios de Castries, en donde se planeaba efectuar una expansión de las líneas. Como el censo había sido levantado en tres niveles geográficos (provincia, ciudad y distrito de enumeración), el nivel menor no era lo suficientemente pequeño como para definir con exactitud el área de interés.

En Chile (cerca de 12 millones de habitantes), el sistema es utilizado tanto por el Instituto Nacional de Estadística (INE) como por el CELADE, siempre con un enfoque dirigido al procesamiento de datos de áreas pequeñas. Un ejemplo característico es el estudio que se efectúa en CELADE sobre las poblaciones indígenas de la comuna de Temuco, donde en primer término se identificaron cartográficamente las áreas pertinentes, para luego conectarlas lógicamente en el REDATAM, estructurando así el universo de estudio.

En Guyana, el sistema se usó para cargar los datos de la encuesta retrospectiva de demografía (GUYREDEM, Retrospective Demographic Survey) de 1986, que a pesar de no ser muy significativa en términos del volumen de información, sirvió para poner a prueba el sistema con archivos ponderados (factor de expansión). La gran ventaja en este caso es la velocidad de procesamiento y el espacio ocupado por los archivos, en comparación con la utilización del SPSS/PC (Statistical Package for the Social Sciences, versión para micros). La desventaja es que el SPSS es mucho más flexible en materia de

REDATAM: RELEVANCIA PARA LOS CENSOS DE 1990

presentación tabular que la actual versión de REDATAM, en tanto entrega los resultados listos para publicación y posee mayores niveles de sofisticación en el procesamiento estadístico (REDATAM cuenta con una interfaz directa con SPSS para estos casos).

El ejemplo de Guyana es doblemente interesante, porque presentó algunos problemas de identificación en los cuestionarios, los que estaban adscritos a Distritos de Enumeración que no habían sido seleccionados para la muestra y por lo tanto no aparecían en los archivos geográficos. Estas divergencias entre la identificación de los cuestionarios y los archivos de control pueden traducirse en una mala ubicación geográfica de los registros censales, lo que constituye una dificultad grave para los propósitos del usuario de áreas pequeñas.

Otro posible elemento de interferencia lo constituyen los cambios de estructura geográfica en los períodos intercensales, como sucedió en los casos de Chile y Guyana, los que probablemente perjudicarán la comparabilidad histórica de la información censal.

IV. CONCLUSIONES.

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 actuales los conceptos y la experiencia adquirida deberían incorporarse a las decisiones sobre planificación y organización, las que a la larga afectan la posterior utilización de los datos para áreas pequeñas.

En primer lugar, una buena cartografía es imprescindible para la identificación correcta de los cuestionarios dentro de los niveles geográficos menores. Una base de este tipo que pueda actualizarse en los períodos intercensales, resulta esencial para la explotación masiva de los datos. Una sugerencia en este sentido sería mantener un sistema computarizado de almacenamiento de las cartas, que proporcionaría una garantía de cobertura

REDATAM: RELEVANCIA PARA LOS CENSOS DE 1990

censal y una producción automática de mapas con mayor precisión y rapidez, permitiendo en el futuro, una conjugación de los datos censales con los cartográficos, a través de la presentación gráfica de los resultados. ⁴

En el censo de Colombia, en 1985, se utilizó un sistema de cartografía computarizada, desarrollado internamente en el Departamento Administrativo Nacional de Estadística (DANE), denominado Micro-Map, para microcomputadores Apple IIe. ⁵

La definición previa de los niveles mínimos de identificación geográfica es central, ya que, a la larga, limitará la posibilidad del sistema de obtener información a niveles menores. Por ejemplo, si la identificación llega hasta la manzana, no será posible obtener datos a nivel de face de manzana. Esta definición debe determinarse en las primeras etapas de la planificación, ya que afectará la cartografía, el diseño de cuestionarios, el operativo censal, etc.

El Control Primario de las boletas censales también juega un papel importante en los resultados a nivel de áreas pequeñas. Este proceso, ejecutado luego del ingreso de los datos, es necesario para verificar, entre otros aspectos, la exactitud en las variables de Ubicación Geográfica de los cuestionarios, lo que permite asegurar la correcta distribución de los mismos dentro de los niveles geográficos menores codificados (sector, manzana, etc.).

Los beneficios de una buena cartografía y un control primario adecuado no parecen necesarios ni para el total del país ni para una región o departamento, en la medida en que la población total no se ve afectada. En cambio, un error al identificar un área pequeña a nivel de barrio o manzana

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, Notas de Población No.41, Santiago, Agosto de 1986.

⁵ Ferro Calvo, M. El Proyecto de la Cartografía Censal, Taller de Análisis y Evaluación de los Censos de Población, págs. 369-400, Buenos Aires, Mayo de 1985.

REDATAM: RELEVANCIA PARA LOS CENSOS DE 1990

puede ser funesto para los futuros usuarios en la obtención de información a ése nivel geográfico.

La comparabilidad histórica entre un censo y otro debe mantenerse a cualquier costo, no sólo en lo que se refiere a las preguntas básicas de los cuestionarios, sino también en lo que dice relación con las subdivisiones geográficas. El concepto de áreas mínimas de comparación debe ser rigurosamente observado, es decir, la identificación geográfica mínima con la cual se pueden comparar los resultados censales debe ser realmente la menor posible, a fin de posibilitar el estudio comparativo para áreas pequeñas.

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.

(docsem03/celev2)

