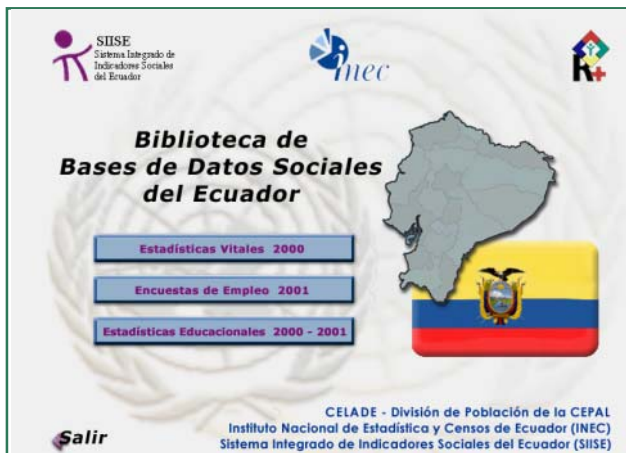


una publicación del CELADE - CEPAL

LC/L.1812



Pantalla principal de acceso a la Biblioteca de Datos Sociales (BBDS) del Ecuador, una aplicación hecha con R+xPlan y con R+WebServer (Figura 2.1 del artículo sobre la BBDS que comienza en la página 7).

## Tabla de contenido

Editorial... 1
Una metodología para estimar el déficit habitacional usando datos censales .... 2
Ecuador: Diseño de una Biblioteca de Bases de Datos Sociales (BBDS) .... 7
Nota Técnica sobre R+G4: Nuevas Funciones para el Manejo de Bases de Datos .... 9
Sistema de Indicadores para el Seguimiento de la Conferencia Internacional sobre la Población y el Desarrollo y de la Plataforma de Acción de Beijing .... 10
The REDATAM Software for Census Data Dissemination and Analysis ... 12
Lista de microdatos censales disponibles / List of available census microdata ... 19

Este documento no ha sido sometido a revisión editorial.

## Editorial

La disseminación y uso de los datos censales de población y vivienda de la ronda de 2000 en Latinoamérica y el Caribe sigue siendo objetivo central del desarrollo de la familia Redatam (véanse las págs.12-18). Varias Oficinas Nacionales de Estadística (ONE) están poniendo sus datos disponibles mediante aplicaciones creadas con R+xPlan (véase *Redatam Informa* 2001 para detalles técnicos). Algunos optaron por hacer sus microdatos más ampliamente accesibles mediante R+WebServer, que maneja aplicaciones de Internet (véanse págs. 17-18 de este informativo)

Para ilustrar cómo los datos censales de 2000 pueden usarse para cuantificar un problema social, el primer artículo describe una metodología para estimar déficit de viviendas. El segundo muestra el diseño de la Biblioteca de Bases de Datos Sociales (BBDS) de Ecuador, que contiene diversas bases de datos Redatam y el tercero muestra los esfuerzos regionales para desarrollar indicadores que monitoreen las metas de conferencias internacionales. El cuarto artículo revisa la situación actual del software Redatam, con ejemplos y sitios web regionales y nacionales. Una nota técnica describe las facilidades para crear nuevas bases.

Reconociendo que los datos de la ronda censal 2000 pueden tener más valor comparados con los anteriores, CELADE entrega un listado de microdatos censales en su poder para 20 países latinoamericanos desde la ronda de 1960, muchos de ellos recuperados mediante un proyecto con la Universidad de Minnesota.



Manténgase Al Día todos los días  
[www.eclac.cl/celade/redatam/](http://www.eclac.cl/celade/redatam/)

## Una metodología para estimar el déficit habitacional usando datos censales

Jorge Rodríguez (jrodriguez@eclac.cl)  
 Sebastián Carrasco (scarrasco@eclac.cl)  
 Celade-Cepal, Santiago, Chile

Para estimar, localizar y calificar el déficit habitacional, el censo tiene potencialidades enormes, porque: (a) proporciona información desagregada hasta el menor nivel geográfico que permite el secreto estadístico; y (b) recoge información de al menos dos de las entidades relevantes para su análisis, vivienda y persona; más aun, en la mayoría de los censos de la región obtiene información de una tercera entidad relevante: el hogar (Rodríguez, 1999).

Considerando lo anterior y teniendo en cuenta que la información censal ha sido —pese a los avances en materia de capacidad y velocidad para procesar grandes bases de datos (como las censales) y a las tendencias de mayor apertura en el acceso a dichas bases de datos— poco aprovechada en el campo de vivienda, CELADE elaboró, a fines del decenio pasado, un procedimiento destinado a estimar el déficit habitacional haciendo uso de la información censal. Dos principios generales guiaron la construcción del procedimiento. El primero fue el de *simplicidad* tanto para permitir un uso masivo del mismo como para facilitar la adaptación y réplica a los diferentes casos nacionales; tal principio se tradujo en una concentración en las viviendas particulares ocupadas y en los requerimientos habitacionales básicos de primera vivienda. El segundo fue el de *innovación* tanto en lo que atañe al uso de Redatam para el procesamiento de la información censal como en lo que respecta a la lógica y las categorías tradicionalmente usadas para la estimación del déficit habitacional (MIDEPLAN, 2001; CELADE, 1996); tal principio se expresó en la incorporación del núcleo familiar como unidad básica del requerimiento habitacional, en la producción de una *matriz integrada de resultados* que combina déficit habitacional cuantitativo y déficit habitacional cualitativo (figura 1.1) y en la obtención de resultados para esta matriz de tres entidades relevantes (viviendas, hogares y personas), aprovechando la versatilidad de Redatam para el manejo de información jerárquica.

El procedimiento, que fue aplicado al censo de población y vivienda de Chile de 1992 (las ilustraciones gráficas corresponden a dicha aplicación), ha sido adaptado a otros censos de la región, como en Perú 1993, Ecuador 2000 y Bolivia 2001. A grandes rasgos, el procedimiento puede describirse de acuerdo a las siguientes etapas:

Figura 1.1. Matriz integrada de déficit habitacional: Tipos de déficit (variable Carencia) por Falta de vivienda (variable Allega).

TIPOS DE DÉFICIT	POR FALTA DE VIVIENDA (cuantitativo): VARIABLE ALLEGA			
	Viviendas sin allegamiento	Viviendas sólo con allegamiento externo	Viviendas sólo con allegamiento interno	Viviendas con ambos tipos de allegamiento
Viviendas sin deficiencias	1	2	3	4
Viviendas con deficiencias recuperables	5	6	7	8
Viviendas con deficiencias irrecuperables	9	10	11	12

Fuente: Celade, 1996.

Figura 1.2. Interpretación habitacional de cada celda de la matriz de Figura 1.1 (véase un ejemplo empírico en Fig 1.5).

CONDICIÓN HABITACIONAL	CELDA DE LA MATRIZ
Sin problemas habitacionales	1
No allegado y deficiencias cualitativas recuperables (sólo déficit cualitativo)	5
No allegado y deficiencias cualitativas irrecuperables (sólo déficit cuantitativo por mala calidad)	9
Allegamiento sólo externo y sin deficiencias cualitativas (sólo déficit cuantitativo por allegamiento)	2
Allegamiento sólo externo y deficiencias cualitativas recuperables	6
Allegamiento sólo externo y deficiencias cualitativas irrecuperables	10
Allegamiento sólo interno y sin deficiencias cualitativas	3
Allegamiento sólo interno y deficiencias cualitativas recuperables	7
Allegamiento sólo interno y deficiencias cualitativas irrecuperables	11
Ambos tipos de allegamiento y sin deficiencias cualitativas	4
Ambos tipos de allegamiento y deficiencias cualitativas recuperables	8
Ambos tipos de allegamiento y deficiencias cualitativas irrecuperables	12

Fuente: Celade, 1996.

1) Examinar —revisando los formularios censales y efectuando procesamientos de prueba de la base de datos con Redatam— la información en la base de datos con los siguientes propósitos:

- i) identificar y analizar las variables que permiten delimitar el universo de estudio, que corresponde a las viviendas particulares ocupadas con moradores presentes (variables tipo y condición de ocupación de la vivienda). En el caso de Chile, el 92% de las viviendas censadas clasificó en tal condición.
- ii) identificar y analizar las consultas que permiten medir aspectos cualitativos de la vivienda y por esa vía detectar situaciones de déficit habitacional cualitativo. En el caso de Chile se consideraron consultas relacionadas con las siguientes dimensiones: (1) la materialidad de la vivienda, es decir, paredes exteriores, techo y piso, cuyas categorías de respuesta se clasificaron, para todo

## ¿Que es Redatam?

**Redatam** es acrónimo de **RE**cuperación de **DAT**os para **Á**reas pequeñas por **MI**crocomputador. *Redatam+G4* (abreviatura: *R+G4*), la cuarta y más reciente generación del software (2001), puede ser usada en español, inglés o portugués, con Windows 95, 98, NT4, 2000 y XP, en cualquier microcomputador compatible con IBM. Este recuadro describe las capacidades básicas del software; el recuadro *Productos de la familia Redatam* en la página 4 indica los propósitos específicos de cada uno de los productos.

Redatam utiliza una base de datos comprimida, que contiene *microdatos* y/o información agregada con millones de registros de personas, viviendas, manzanas de ciudades o cualquier división administrativa de un país. Esos datos pueden provenir de cualquier combinación de censos, encuestas u otras fuentes. Se puede procesar una base de datos en asociación con bases de datos *externas* como dBASE.

Un usuario con una base de datos Redatam puede definir cualquier área geográfica de interés (desde manzanas de una ciudad) o combinaciones de esas áreas, crear nuevas variables y producir tabulados rápidamente. Los datos de diferentes niveles geográficos pueden ser combinados jerárquicamente para crear variables agregadas, y los resultados pueden desplegarse sobre mapas en Redatam mismo o transferirse a un Sistema de Información Geográfico (SIG).

La primera generación de Redatam, para DOS, se lanzó en 1987. Casi todos los países de América Latina y el Caribe tienen bases de datos de un o más censos y hay usuarios del software en África y Asia y en varios países desarrollados. Todas las generaciones de Redatam han sido desarrolladas por el Centro Latinoamericano y Caribeño de Demografía (CELADE), División de Población de la CEPAL. Se puede descargar la 4ª generación, *R+G4*, sin costo alguno, desde <http://www.eclac.cl/celade/redatam>.

el país, en no deficitarias, deficitarias recuperables y deficitarias no recuperables de acuerdo con la realidad habitacional y ambiental de Chile; (2) el espacio físico, es decir, el hacinamiento, que se midió a escala de vivienda como número de residentes por dormitorio, y cuyos resultados se clasificaron en sin hacinamiento (menos de 2.5) con hacinamiento recuperable (entre 2.5 y 5) y con hacinamiento irrecuperable, más de 5; (3) los servicios, es decir, alumbrado eléctrico, agua potable y servicio higiénico, cuyas categorías de respuesta se clasificaron, para todo el país, en no deficitarias y deficitarias recuperables de acuerdo con la realidad social de Chile.

iii) identificar y clasificar las consultas que permiten medir la falta de vivienda y por esa vía detectar el componente tradicional del déficit cuantitativo. Típicamente se considera que los hogares son las unidades que requieren vivienda, por

lo que cabe asegurarse de que el censo distinga entre vivienda y hogar, pues en algunos pocos censos de la región tal distinción se omite (Rodríguez, 1999). Pero el procedimiento propuesto no concluye en los hogares, pues procura extenderse hacia la unidad que, en rigor, requiere vivienda, cual es el núcleo familiar. Y esto último es un desafío mayor, pues tal unidad no existe como entidad (es decir, una unidad respecto de la cual se recoge información primaria y que se expresa como un nivel jerárquico en la base de datos) en los censos de la región. Sólo algunas encuestas de hogares —por ejemplo la Encuesta de Caracterización Socio Económica Nacional (CASEN) en Chile ([www.mideplan.cl](http://www.mideplan.cl)) y la Encuesta Nacional de Hogares (ENAHOG) en Perú ([www.inei.gov.pe](http://www.inei.gov.pe))— identifican ex-ante al núcleo familiar, lo que permite la obtención de información referida específicamente al núcleo familiar.

En el caso del censo, es posible imputar la existencia de núcleos familiares dentro de los hogares considerando algunos supuestos y convenciones (una parte de ellas esencialmente arbitrarias) y el uso combinado de varias consultas. La principal de ellas es la pregunta por parentesco del jefe de hogar (el servicio doméstico se excluye por definición). Pero también se utilizan las preguntas por edad, estado civil, paridez total y de hijos sobrevivientes. Con ellas, el procedimiento identifica: (a) núcleo principal, (b) núcleos de hijos/as del jefe; (c) núcleos de hermanos del jefe; (d) núcleos de padres del jefe; (e) núcleos de otros parientes del jefe, (f) núcleos de no parientes del jefe. Explicar los algoritmos, supuestos y convenciones necesarias para identificar estos núcleos supera los objetivos de esta nota; en todo caso, están detallados en CELADE, 1996. Las revisiones posteriores del procedimiento su-

Figura 1.3: Aplicación del procedimiento: resultados para una comuna pequeña de Chile a escala de vivienda.

TIPOS DE DÉFICIT	POR FALTA DE VIVIENDA (cuantitativo): VARIABLE ALLEGA				
	Sin allegamiento	Sólo con allegamiento externo	Sólo con allegamiento interno	Con ambos tipos de allegamiento	Total
Sin deficiencias	14 506	62	1 016	13	15 597
Deficiencias recuperables	2 124	15	169	5	2 313
Deficiencias irrecuperables	111	1	14	1	127
<b>Total</b>	<b>16 741</b>	<b>78</b>	<b>1 199</b>	<b>19</b>	<b>18 037</b>

Fuente: Procesamiento especial de la base de datos del censo de Chile, 1992.

gieren que los principales problemas teóricos, metodológicos y de programación en este terreno se encuentran en la identificación de núcleos de hijos/as del jefe de hogar. En los otros casos, los eventuales problemas atañen más bien a las convenciones usadas para contabilizar núcleos. Cabe subrayar que el procedimiento sólo logra imputar un valor de número de núcleos por hogar pero no permite tener información específica sobre los núcleos, lo que impide caracterizar a los núcleos afectados por el déficit.

2) Una vez efectuadas las anteriores tareas, se usa Redatam para generar variables a escala de vivienda. Así, las viviendas del universo adquieren la condición de no deficitaria cualitativamente (valor 0), deficitaria recuperable (valor 1) y deficitaria irrecuperable (valor 10) en cada rubro de calidad habitacional. Luego, se genera un índice de calidad de la vivienda que corresponde a una suma simple de todas las variables antes creadas. Índice con valor cero significa ausencia de déficit cualitativo. Las viviendas con un índice entre 1 y 8 conforman el déficit cualitativo. Las viviendas con un índice 10 ó superior tienen alguna deficiencia irrecuperable y por ende engrasan el déficit cuantitativo, pues la intervención habitacional que ameritan es una solución total, una nueva vivienda. De esta manera, con una programación en Redatam que no reviste complejidad —sintaxis básica de comando DEFINE con AS RECODE y AS SUM—, se genera la primera variable de la matriz integrada del déficit habitacional, el principal resultado de este procedimiento. La segunda variable de la matriz está vinculada a la cantidad de unidades que requieren vivienda. Siguiendo una expresión usada en algunos países de la región (allegamiento), esta variable tiene cuatro categorías: (a) sin allegamiento: un solo hogar y núcleo familiar en la vivienda; (b) con allegamiento externo: más de un hogar por vivienda pero cada hogar contiene sólo un núcleo familiar; (c) con allegamiento interno: sólo un hogar en la

Figura 1.4: Aplicación del procedimiento: resultados para una comuna pequeña de Chile a escala de hogares.

TIPOS DE DÉFICIT	POR FALTA DE VIVIENDA (cuantitativo): VARIABLE ALLEGA				Total
	Sin allegamiento	Sólo con allegamiento externo	Sólo con allegamiento interno	Con ambos tipos de allegamiento	
Sin deficiencias	14 506	128	1 016	29	15 679
Deficiencias recuperables	2 124	31	169	10	2 334
Deficiencias irrecuperables	111	2	14	2	129
<b>Total</b>	<b>16 741</b>	<b>161</b>	<b>1 199</b>	<b>41</b>	<b>18 142</b>

Fuente: Procesamiento especial de la base de datos del censo de Chile, 1992.

## Productos de la familia Redatam

**Redatam+G4** (abreviado *R+G4*), la cuarta y más nueva generación del software Redatam —introducido en 2000—, cuenta con los siguientes miembros:

- **R+Process**, para el procesamiento de datos —provenientes de áreas geográficas seleccionadas por el usuario— con poderosos programas escritos en el lenguaje de comandos Redatam o para procesar con el restringido Asistente que no necesita ninguna habilidad en programación y tampoco conocimiento del lenguaje Redatam. Este módulo incluye, además, facilidades para expandir, agregar, conectar, descargar y otras acciones, las actuales bases de datos *R+G4*.
- **R+Create**, que permite crear bases de datos jerárquicas en un formato propio de *R+G4* a partir de archivos en formato ASCII o xBase. Los archivos ASCII pueden tener sus registros (layouts) definidos en IPS, ISSA o CHILLAN (formato interno de *R+G4*).
- **R+xPlan** para construir aplicaciones que a los usuarios finales permitan acceder, sin ningún conocimiento de Redatam, a microdatos protegidos con el fin de obtener indicadores que con frecuencia implican cálculos complejos.
- **R+WebServer**, el más reciente (2002) integrante de la familia, con el cual se puede solicitar información en línea de una base de datos *R+G4* a través de intranet, extranet o Internet. Al igual que en el caso de los otros miembros de la familia Redatam, el *R+WebServer* se funda en la maquinaria estadística Redatam, que posee una alta eficiencia en un ambiente en línea.
- **Herramientas de apoyo espacial para la toma de decisiones**, desarrolladas en la Faculty of Environmental Studies, University of Waterloo, Ontario, Canada N2L 3G1.

**AccessPlan**: atención primaria en salud y planificación familiar; **EduPlan**: planificación educacional, y **TourPlan**: planificar el desarrollo turístico. Todas pueden usar los datos procesados con *Redatam+G4* (o con *winR+*). Véase: [www.fes.uwaterloo.ca/Tools/](http://www.fes.uwaterloo.ca/Tools/); contacto: [gbhall@fes.uwaterloo.ca](mailto:gbhall@fes.uwaterloo.ca).

Los tres primeros miembros de la familia Redatam forman un solo conjunto de módulos del programa *R+G4*; trabaja con Windows 95, 98, NT4.0, 2000 y XP. *R+WebServer* opera con un servidor Windows. Si se opera con Windows 3.x, tiene que usarse la versión anterior, **winR+ v1.2**.

**R+G4** puede descargarse, sin costo alguno, del sitio del CELADE: <http://www.eclac.cl/celade/redatam>. El

vivienda pero que contiene más de un núcleo familiar; (d) con ambos tipos de allegamiento: hay más de un hogar y en al menos uno de ellos hay más de un núcleo.

3) Con estas dos variables se obtiene la matriz integrada del déficit habitacional (figura 1.1). Se denomina “integrada” porque cruza el déficit cualitativo (variable calidad de la vivienda) con el déficit cuantitativo (variable allegamiento). El resultado es una matriz con 12 casilleros en que cada uno representa una si-

Figura 1.5. Aplicación del procedimiento: resultados para una comuna pequeña de Chile a escala de núcleos familiares.

DÉFICIT	NÚMERO DE NÚCLEOS				Total
	1 Núcleo	2 Núcleos	3 Núcleos	4 Núcleos	
Sin problemas habitacionales	14 506	-	-	-	14 506
No allegado y deficiencias cualitativas recuperables (sólo déficit cualitativo)	2 124	-	-	-	2 124
No allegado y déficit cual. irrecuperables (sólo déficit cuantitativo por mala calidad)	111	-	-	-	111
Allegamiento sólo externo y sin defíc cual. (sólo déficit cuantitativo por allegamiento)	128	-	-	-	128
Allegamiento sólo externo y deficiencias cualitativas recuperables	31	-	-	-	31
Allegamiento sólo externo y deficiencias cualitativas irrecuperables	2	-	-	-	2
Alleg. sólo interno y sin defíc cualitativas	-	934 (1 868)	72 (216)	10 (40)	1 016 (2 214)
Allegamiento sólo interno y deficiencias cualitativas recuperables	-	149 (298)	19 (57)	1 (4)	169 (359)
Allegamiento sólo interno y deficiencias cualitativas irrecuperables	-	5 (10)	9 (27)	-	14 (37)
Ambos tipos de allegamiento y sin deficiencias cualitativas	15	13 (26)	1 (3)	-	29 (44)
Ambos tipos de allegamiento y deficiencias cualitativas recuperables	5	5 (10)	-	-	10 (15)
Ambos tipos de allegamiento y deficiencias cualitativas irrecuperables	1	1 (2)	-	-	2 (3)
<b>Total Hogares</b>	<b>16 923</b>	<b>1 107</b>	<b>101</b>	<b>11</b>	<b>18 142</b>
<b>Total núcleos</b>	<b>16 923</b>	<b>2 214</b>	<b>303</b>	<b>44</b>	<b>19 484</b>

Fuente: Procesamiento especial de la base de datos del censo de Chile, 1992.

tuación específica de déficit habitacional a escala de vivienda (figura 1.1). Sólo en un casillero (el extremo superior izquierdo en el diagrama) se encuentran las viviendas sin déficit, esto es, que sólo tienen un hogar y un núcleo familiar y que no tienen deficiencias cualitativas (materiales, espacio y servicios adecuados). Los restantes 11 casilleros clasifican viviendas con algún tipo de déficit. La fila inferior así como la última columna pueden considerarse, en principio, como situaciones particularmente graves. En el primer caso se trata de viviendas de calidad irrecuperable, a causa de lo cual ameritan soluciones totales (nuevas viviendas). En el segundo caso, se trata de viviendas donde se verifican simultáneamente ambos tipos de allegamiento, sugiriendo estrecheces, incomodidades y conflictos cotidianos más marcados o intensos.

4) Esta matriz está definida a escala de vivienda, por lo que no es suficiente para hacer las estimaciones de déficit habitacional. Esto porque, por ejemplo, indica que hay una cierta cantidad de viviendas irrecuperables pero no proporciona antecedentes sobre la cantidad de hogares y núcleos familiares que habita en ellas, cifra que define la cantidad de nuevas viviendas que habría que construir para enfrentar ese tipo de déficit. Así las cosas, para proceder a la estimación del déficit habitacional es necesario replicar la matriz antes descrita a escala de hogares y de núcleos familiares. Replicarla a escala de hogares no

reviste mayor problema, pues Redatam permite varias formas de hacerlo. La recomendada es usar la opción WEIGHT y aplicarla a una variable a escala de vivienda que corresponda al número de hogares en la vivienda. Si esta no forma parte de la base de datos original, cabe crearla con un DEFINE (a escala de vivienda) y un COUNT (de hogares por vivienda) para luego grabarla utilizando la instrucción SAVE. Replicar la matriz a escala de núcleo familiar es imposible, pues, como ya se explicó, aquél no es una unidad de recolección de información del censo y tampoco una entidad en la base de datos Redatam. Sin embargo, sí puede obtenerse la matriz según núcleos bajo cada categoría de déficit usando la opción WEIGHT sobre una variable nueva (y grabada con SAVE) que contenga el número de núcleos de la vivienda.

5) Posteriormente, la matriz se replica a escala de persona, usando la opción WEIGHT para una variable construida *ad hoc* a escala de vivienda que contenga la cantidad de personas por vivienda (en principio, excluyendo servicio doméstico), lo que permite tener una cuantificación de la población afectada por las distintas situaciones (casilleros) de déficit habitacional.

6) Luego se aplican algunos algoritmos simples (CELADE, 1996) para calcular el déficit. En lo medular, estos algoritmos permiten distinguir entre núcleos familiares que habitan una misma vivienda (y por tanto se hallan en un mismo casillero de la matriz) pero que tienen distintas situaciones de déficit habitacional. Un

ejemplo ilustrativo se presenta en las figuras 1.3 y 1.4 con la situación habitacional “allegamiento sólo externo y deficiencias cualitativas recuperables” (celda 6 de la matriz, figura 1.1). En 1992, en la comuna chilena de Lampa había 15 viviendas (figura 1.3) y 31 hogares (figura 1.4) en dicha celda. El cómputo del déficit que se deriva de dicha celda es, por tanto, 16 viviendas de déficit cuantitativo (31-15) y 15 viviendas de déficit cualitativo, es decir que requieren de algún tipo de reparación. Cabe subrayar que el tipo de reparación requerido también podría ser identificado con ulteriores procesamientos de la bases de datos. Un ejemplo algo más complicado se deduce del cotejo de los figuras 1.3, 1.4 y 1.5. La celda 7 tiene 169 viviendas y los mismos 169 hogares; una estimación clásica del déficit habitacional sumaría 169 viviendas con déficit cualitativo. Sin embargo, el procedimiento elaborado por CELADE penetra en los hogares y contabiliza núcleos familiares que se supo-

nen las unidades que efectivamente requieren vivienda. Esto se aprecia en el figura 1.5 donde la celda 7 suma 359 núcleos familiares. De esta manera el aporte de esta celda al déficit sería de 190 viviendas de déficit cuantitativo (359-169) y 169 viviendas de déficit cualitativo. Por cierto, la estimación es altamente sensible a la definición de núcleo familiar. Esta última no es trivial y es, dentro de ciertos límites, un asunto de convención. El espacio disponible no permite ahondar en este tema aunque confiamos poder hacerlo en el futuro en esta misma publicación.

7) Finalmente, cabe reiterar que la versatilidad de Redatam para análisis jerárquicos hace posible efectuar caracterizaciones no sólo de las viviendas, como se expuso en el acápite previo, sino también de los ho-

## Talleres sobre Redatam

Con la finalidad de seguir apoyando a las instituciones nacionales en el aprovechamiento de la información socio-demográfica y sectorial para el diseño y ejecución de programas y políticas sociales, el CELADE organiza con frecuencia talleres destinados a capacitar a profesionales y funcionarios en el manejo de *Redatam+G4*.

Durante el año 2003 se realizarán los siguientes talleres, repitiéndolos según la necesidad.

**Taller básico:** *Redatam+G4* como una Herramienta de Apoyo a la Toma de Decisiones.

**Taller avanzado:** Generación de Bases de Datos y Procesamiento de Indicadores Sociodemográficos con *Redatam+G4*.

**Taller de especialización:** Creación de bases de datos, construcción de aplicaciones *R+xPlan* y/o *R+G4 Web Server*.

Los talleres se realizarán en la CEPAL, Santiago; cada uno tendrá una duración de entre 5 y 20 días en horario completo. Se puede realizar en otro lugar a pedido especial.

Los participantes (o las instituciones que los patrocinan) deberán costear su pasaje y estadía en Santiago (monto de referencia: US \$100 por día).

Para un taller de 5 días se deberá cancelar, por concepto de matrícula, la suma de: US\$250 (dólares o su equivalente en \$ chilenos) para latinoamericanos y caribeños que sean funcionarios de oficinas públicas, universidades u ONG y US\$400 para otros participantes. Los costos de los otros talleres dependerán de su duración.

Véase [www.eclac.cl/celade/](http://www.eclac.cl/celade/) para fechas, información adicional y el formulario de inscripción.

## Bibliografía

Rodríguez, Jorge (1999), Información censal relevante para la medición del déficit habitacional, *Notas de Población*, No. 69, páginas 91-58.

Ministerio de Planificación y Cooperación de Chile (MIDEPLAN) (septiembre 2001), *Situación habitacional 2000. Informe Ejecutivo. Resultados Preliminares*, Santiago, MIDEPLAN, [www.mideplan.cl/sitio/Sitio/estudios/documentos/informe vivienda2000.pdf](http://www.mideplan.cl/sitio/Sitio/estudios/documentos/informe vivienda2000.pdf)

CELADE (1996), *Déficit habitacional y datos censales sociodemográficos: una metodología*, Santiago, CEPAL/CELADE, LC/DEM/R.267, Serie B, N° 114.

Encuesta de Caracterización Socio Económica Nacional (CASEN) en Chile ([www.mideplan.cl](http://www.mideplan.cl)).

Encuesta Nacional de Hogares (ENAH) en Perú ([www.INEI.gob.pe](http://www.INEI.gob.pe)). ☒



## Publicaciones Periódicas del Celade

[www.eclac.cl/celade/publicaciones](http://www.eclac.cl/celade/publicaciones)



## Introducción<sup>2</sup>

La accesibilidad a la información adecuada, actualizada y pronta fomenta la correcta identificación de proyectos sociales, la adecuada caracterización de sus beneficiarios y la elaboración actualizada de diagnósticos. El acceso rápido y fácil a esta información es el desafío de todas las nuevas herramientas computacionales de hoy. En este sentido, si se cuenta con una biblioteca de información social a nivel nacional y que, a la vez, permita su desagregación geográfica da la posibilidad de establecer prioridades basadas en jerarquizaciones y se transforma en una herramienta de suma utilidad.

En un convenio entre el Banco Interamericano (BID) y el Gobierno de Ecuador se planteó desarrollar una *Biblioteca de Bases de Datos Sociales* (BBDS) generada a partir del Sistema de Estadísticas Nacionales (SEN). Su objetivo general es promover el uso de la información y bases de datos del SEN mediante mecanismos de amplia difusión y usando las nuevas tecnologías de la información, el desarrollo de instrumentos para el análisis aplicado a las necesidades de las políticas y la formación de recursos humanos capaces de usar la información en su práctica profesional. La Biblioteca se visualiza como un sistema de información permanentemente actualizado que integre en una sola plataforma de fácil uso y acceso, los indicadores sociales, de población y de desarrollo existentes en Ecuador.

Se decidió hacer primero un prototipo de la BBDS, utilizando los productos de la familia Redatam. Como medios de difusión se definió: a) un CD que integra en una aplicación el programa de manejo estadístico (*R+xPlan*), interfase y base de datos en un solo sistema; b) procesamiento con *R+Process* de bases de datos en un servidor de una LAN (Red de Área Local); y, por último, c) con procesamiento a distancia por Intranet o Internet usando *R+Web-Server* para acceder, analizar y procesar la información de la biblioteca de datos.

En esta nota se esboza el diseño y desarrollo del prototipo; por su importancia en el diseño de un sistema, en un anexo se presentan las estructuras jerárquicas de los diferentes tipos de datos incluidos en este prototipo.

## Implementación del prototipo

El objetivo del proyecto era producir un prototipo de sistema computacional que sirva de base para conformar la BBDS que integre diversas fuentes de información derivadas del SEN. El prototipo debía proporcionar:

- *Base de datos*: Versión electrónica de los indicadores, para que los usuarios consulten y puedan obtener e integrar resultados en sus

aplicaciones y documentos;

- *Meta-información*: Descripción detallada de cada indicador, incluidas las fuentes de información y la metodología empleada en su construcción;
- *Cuadros estadísticos*: Estadística descriptiva sobre cada indicador;
- *Gráficas*: Representación gráfica de las estadísticas correspondientes a cada indicador mediante el uso de diversos gráficos;
- *Mapas temáticos*: Mapas, basados en cartografía digital, que desplieguen las estadísticas correspondientes a cada indicador a determinado nivel geográfico.

Con financiamiento del Convenio BID-Gobierno del Ecuador (No. ATN/JF-6366-EC), el CELADE realizó el prototipo con la ayuda de personal del INEC; en primera instancia se trabajó en Santiago durante dos semanas para completar las siguientes etapas:

- Selección de bases de datos y limpieza de archivos.
- Creación de bases de datos en formato Redatam con *R+Create*.
- Definición de los indicadores que se incluirán en la aplicación.
- Creación de la aplicación prototipo para BBDS usando *R+xPlan*.
- Traspaso de la aplicación a Internet mediante *R+WebServer*.

Luego se continuó trabajando en Ecuador. En conjunto con el INEC se instaló una versión del prototipo de la BBDS en la red local de INEC y se terminaron las aplicaciones *R+xPlan* para difundir en CD con:

- a) contenidos censales de 1982, 1990 y una muestra del 2001.
- b) información de estadísticas vitales, encuestas de empleo y de educación del año 2000.

## Información que se incluirá en la BBDS

Las bases de datos que conformarán la BBDS completa son:

- Censos de población y vivienda,
- Otros censos disponibles (de estudiantes, etc.),
- Encuestas de hogares sobre condiciones de vida,
- Encuestas de hogares sobre empleo y desempleo,
- Encuestas de hogares sobre temas demográficos y de salud materno-infantil,
- Otras encuestas sociales disponibles,
- Bases de datos provenientes de registros administrativos,
- Clasificadores (ocupación, actividad económica, cultivos, etc.),
- Códigos de ubicación geográfica.

Las bases de datos deben corresponder a todo el período histórico para el que haya información disponible, dada su importancia para los análisis longitudinales y no sólo para el análisis en un punto en el tiempo.

### Estandarización

Para lograr una adecuada relación entre las bases de datos que conformarán la BBDS, se deberá empezar con un exhaustivo análisis y conocimiento de cada base. Desgraciadamente, casi todas las bases de datos presentan algún grado de problema cuando se trata de compatibilizarla con otro período u otra base de datos. No es raro que una misma base de datos maneje varias versiones a lo largo de un período de tiempo y que, por ende, presente diferencias entre ellas. Es muy común encontrar alguna variable inconsistente o incompleta. La gran mayoría de las bases de datos adolece de una notable falta de estandarización, de errores de captura y algún grado de basura; además de entorpecer muchos de los esfuerzos para integrarlas en un solo sistema, ello hace su manejo aún más difícil. Es necesario estandarizar las distintas bases de datos que conformarán la BBDS y considerar la limpieza, corrección, estandarización, complementación, validación y deduplicación de las bases.

La estandarización de las bases debe ser contemplada en el proceso de organización de la BBDS y estar en manos de un equipo técnico liderado por una persona con experiencia en este tipo de trabajos.

### Documentación

En su versión de aplicación en red local, el prototipo de BBDS presenta documentación en línea sobre:

- Diccionario de cada base de datos, donde se describe cada variable y sus características y todas las características de la estructura de la base de datos.
- Ayuda en línea, que contiene un manual de procedimientos para trabajar la aplicación en línea.

Figura 2.2. Estructura jerárquica de las bases de datos BBDS de Ecuador: Estadísticas vitales.

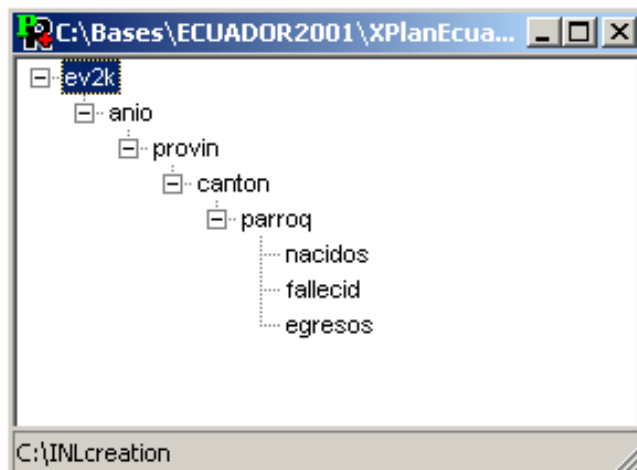


Figura 2.3. Estructura jerárquica de las bases de datos BBDS de Ecuador: Encuesta de empleo.



También incluye manuales del programa Redatam y, además, las funciones de programación para generar indicadores.

- Metadatos; para cada base de datos se incluye una sección descriptiva de la base misma. Bajo esta sección se deberían incluir todas las consideraciones metodológicas sobre la base de datos en particular.

A diferencia de muchos sistemas, el tercer tipo de documentación, es de gran importancia para la funcionalidad de la BBDS. Los metadatos deben organizarse en una base de datos que permita consultar de manera rápida y confiable las características de los datos que componen cada base de datos existente en la BBDS. Se utilizó un sistema comercial [RoboHelp (www.ehelp.com)] para integrar los metadatos como un tipo de “Ayuda” dentro de la BBDS. Esta documentación también debe ayudar a construir una memoria histórica sobre las distintas metodologías, clasificaciones, preguntas y coberturas que se han desarrollado.

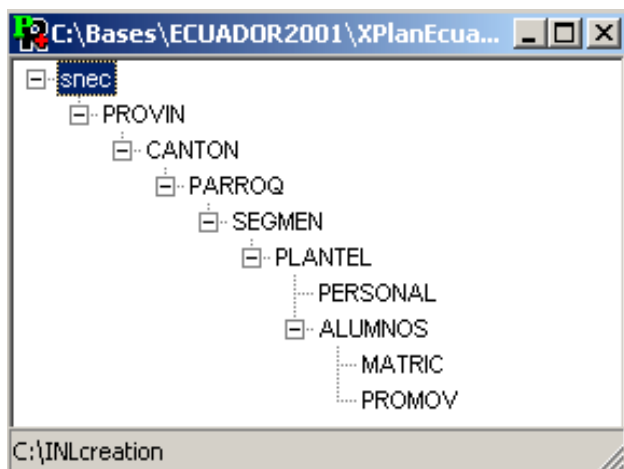
### Funcionamiento de la BBDS

La BBDS está integrada al sistema de difusión de información del INEC, es decir, debe seguir los lineamientos establecidos en la propuesta de políticas de difusión (véase la pantalla principal en figura 2.1, pág. 1). El prototipo se instaló en la red local del INEC, pero está en ligazón con a la página web del INEC. Ello significa que este organismo es quien debe coordinar la creación y mantenimiento de la BBDS.

Se prevé que la BBDS se difunda a través de medios electrónicos como Internet o a través de una red local. También se dispuso la creación de CD con el prototipo de la BBDS. Todos tienen la pantalla principal de acceso a la biblioteca de datos (figura 2.1). Otra forma de difusión es creando un centro de consultas en las oficinas del INEC, probablemente dentro de la biblioteca del Instituto.



Figura 2.4. Estructura jerárquica de las bases de datos BBDS de Ecuador: Educación.



### Anexo: Establecimiento de la jerarquía para cada tipo de datos

La BBDS comprende varias bases de datos individuales de Redatam, y cada una tiene su propia jerarquía. En *estadísticas vitales* se definió una estructura jerárquica que comienza con el año, y se espera tener a futuro una base multi-temporal de datos en donde se podrán obtener tabulados que incluyan varios años. Además, esta jerarquía permite tener tabulados para cada año, provincia, cantón y parroquia de residencia en forma independiente. Como se puede ver en la figura 2.2, para cada parroquia hay datos sobre nacimientos, defunciones y egresos hospitalarios. En el caso de la *encuesta de empleo* (figura 2.3) se definió bajo el hogar una entidad con las variables de empleo y otra con variables del miembro del hogar.

Para la base de datos de *educación* (fig. 3.4), se definió bajo parroquia la entidad segmento que permite identificar las parroquias urbanas y bajo ésta el plantel educacional. Bajo cada plantel, la información se agrupó en dos: una para el personal y otra para los alumnos. Bajo alumnos se definió un grupo como matrícula, que corresponde al inicio del período lectivo y otro grupo que corresponde al término del período lectivo.

### Notas

<sup>1</sup> Este artículo describe parte de la propuesta y su implementación por el CELADE a solicitud del Gobierno de Ecuador en el marco del Convenio BID-Gobierno del Ecuador, No ATN/JF-6366-EC. En la elaboración del prototipo trabajaron: Blanca Almachi, Wilson Ortiz, del INEC-Ecuador; Alejandra Silva y Lenin Aguinaga de Celade-Cepal, Santiago, Chile.

<sup>2</sup> Para más información, véase *Propuesta: Política de difusión del Instituto Nacional de Estadística y Censos del Ecuador: Informe final de consultoría* de Gilberto Moncada Vigo, 20 de octubre, 2000. Frente Social del Gobierno del Ecuador - Instituto Nacional de Estadística y Censos (*Convenio BID-Gobierno del Ecuador, No ATN/JF-6366-EC*). Este documento sirvió como antecedente para el trabajo y para este artículo. ❖

## Nota Técnica sobre R+G4: Nuevas Funciones para el Manejo de Bases de Datos

Ari Silva (arisilva@terra.com.br)  
Rio de Janeiro, Brasil

### Pegar bases múltiples

En el menú de Herramientas del R+G4 hay una nueva función que permite pegar varias bases de datos para generar una base combinada. Un ejemplo: el país generó sus bases de datos (una para cada departamento), y ahora quiere producir una base única para todo el país.

Anteriormente, esa función sólo permitía dos bases a la vez, y con ello el proceso para juntar los departamentos era tedioso, lento, sujeto a errores, etc.

Ahora, al elegir esa nueva función en Herramientas, el sistema despliega una lista para que el usuario ingrese con todos los diccionarios que quiera juntar y, desde luego, elige también el diccionario de salida (la base final). Al ejecutar el proceso, el sistema concatena todas las bases de entrada, en el orden especificado, y genera una base única con todas las informaciones.

Para que esa concatenación funcione, es necesario que todas las bases de entrada tengan la misma jerarquía, es decir, que las entidades sean las mismas. Con relación a las variables, sólo aquellas que pertenezcan a todas las bases de entrada serán concatenadas en la base final.

Esa función fue muy útil, por ejemplo, en el Censo Demográfico de Brasil, para producir la base completa de todo el país a partir de las 27 bases de datos existentes (una para cada unidad federativa).

### Descarga múltiple y muestras de bases

Se está programando un nuevo módulo de R+G4, que permita hacer una descarga múltiple de una base de datos, y/o hacer una muestra de registros de la base.

### ¿Qué es una descarga múltiple?

Supongamos que el Instituto de Estadística va a entregar la base del censo a las comunas o departamentos del país, es decir, cada una de las comunas tendrá su propia base de datos, con los registros que correspondan. Es claro que se podría ejecutar el módulo de descarga existente, pero eso significaría una ejecución de ese módulo ¡para cada comuna! Con el nuevo módulo se ejecuta una sola vez y se especifica el quiebre a nivel de comuna, y el sistema generará una base de salida para cada comuna.

### La muestra

Este módulo también tendrá la función de descargar una muestra de los registros de una base de entrada. Por ejemplo, puede producir una base del país para el 10% de las viviendas. Esta función podrá ser aplicada en conjunto (o no) con la función de descarga múltiple. ❖

## Sistema de Indicadores para el Seguimiento de la Conferencia Internacional sobre la Población y el Desarrollo y de la Plataforma de Acción de Beijing

Susana Schkolnik, (sschkolnik@eclac.cl)  
 Fabiana Del Popolo (fdelpopolo@eclac.cl)  
 Celade, Cepal, Santiago, Chile  
 Enrique Peláez, Consultor, Celade  
 (epelaez@eco.unc.edu.ar)

El Centro Latinoamericano y Caribeño de Demografía (CELADE), División de Población de la Comisión Económica para América Latina y el Caribe (CEPAL) y la Unidad Mujer y Desarrollo de la CEPAL, en estrecha coordinación con el Fondo de Población de las Naciones Unidas (FNUAP), están desarrollando un conjunto de actividades con el fin de responder a mandatos regionales e internacionales para seguir apoyando a los países en la elaboración de sus sistemas de información, cuyos indicadores permitan monitorear y evaluar tanto el Programa de Acción (PA) de la Conferencia Internacional sobre la Población y el Desarrollo (CIPD, El Cairo 1994), como el Plan de Acción Regional, y los compromisos adquiridos en otras cumbres y conferencias internacionales celebradas en la década de 1990.

En el marco de estas actividades, y sumando la experiencia que tanto el CELADE como la Unidad Mujer y Desarrollo vienen acumulando en esta materia, se planteó desarrollar un *Sistema Regional de Indicadores para el Seguimiento del Programa de Acción de la CIPD y de la Plataforma de Acción de Beijing*, como un primer paso para el desarrollo de un sistema integrado que permita monitorear y dar seguimiento a los objetivos de las diversas cumbres y conferencias de las Naciones Unidas.

El desarrollo del Sistema conjunto se inicia a partir de dos sistemas existentes: por un lado el Sistema Regional de Seguimiento de la CIPD desarrollado por el CELADE (www.eclac.cl/celade) y, por otro, el Sistema de Estadísticas de Género de la Unidad Mujer y Desarrollo, en particular los indicadores de seguimiento de Beijing (www.eclac.cl/mujer).

Como punto de partida se hizo una revisión de los aspectos conceptuales y metodológicos de ambos sistemas. Esto implicó la identificación de áreas e indicadores comunes; la revisión de la comparabilidad de la información (períodos de referencia, desagregaciones, fuentes); y finalmente la adopción de criterios y definiciones compatibles.

Para la elección del soporte informático se puso como requisito que éste permitiera la construcción de cuadros, mapas y gráficos según la necesidad del usuario y, a la vez, que permitiera desplegar cuadros

Figura 3.1. Portada de inicio del sistema, con botones que enuncian cada cumbre.



y gráficos pre-construidos. Así, dado que el *R+xPlan* satisfacía los requerimientos básicos, se procedió al diseño del sistema en esta aplicación del Redatam. En la figura 3.1 se ilustra la portada de inicio del sistema y, como se espera darle una mayor proyección al proyecto, aparecen otras dos cumbres internacionales que podrían integrarse al mismo en un futuro.

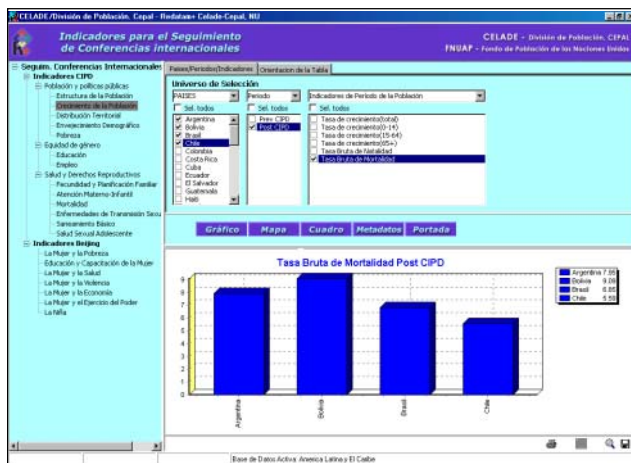
A través de la barra principal se ingresa al listado de áreas prioritarias, capítulos e indicadores de CIPD y Beijing (figura 3.2, izquierda). En esta misma pantalla se pueden obtener cuadros, previa selección de países, períodos e indicadores. También es posible obtener gráficos y mapas (figuras 3.3 y 3.4).

El sistema permite, además, ingresar al procesador de cuadros, gráficos y mapas a través de cada uno de los botones que enuncian cada cumbre (figura 3.1). Al ingresar al sistema mediante cada uno de ellos, se accede a toda la “meta-información” de la respectiva cum-

Figura 3.2. A través de la barra principal se ingresa al listado de indicadores de CIPD y Beijing (izquierda). También se pueden obtener cuadros, previa selección de países, períodos e indicadores.

País	Período	Fecundidad y Planificación Familiar
Argentina	Tasa Global de Fecundidad	3,0
	1985-1990	2,8
	1990-1995	2,6
	1995-2000	2,4
Bolivia	1985-1990	5,0
	1990-1995	4,8
	1995-2000	4,4
	2000-2005	3,9
Brasil	1985-1990	3,0
	1990-1995	2,5
	1995-2000	2,3
	2000-2005	2,1

Figura 3.3. Gráfico que compara la Tasa Bruta de Mortalidad post CIPD en 4 países seleccionados.



bre, como documentos relacionados, tablas pre-estructuradas, perfiles de países, etc. (figura 3.5).

Si bien el acceso a determinados indicadores puede hacerse independientemente para cada conferencia, la información resultante en esos casos es la misma, puesto que se trata de una base de datos común. Lo que puede cambiar es la “mirada” que se da a los resultados obtenidos, en el marco del análisis que desea hacerse.

Por último, tener ambos sistemas integrados evita la duplicación de esfuerzos, y ofrece una complementariedad indiscutible para el seguimiento de las conferencias. Así por ejemplo, hay indicadores de CIPD que son de “contexto” necesarios para Beijing (por ejemplo, los relacionados con la dinámica demográfica) y, al mismo tiempo, indicadores de Beijing relevantes para las áreas temáticas de CIPD (por ejemplo, los relativos a la equidad de género en el empleo).

Figura 3.4. Mapa que compara la Tasa Bruta de Mortalidad post-CIPD en los países de la región.



Con base en el Sistema Regional, el CELADE y la Unidad Mujer y Desarrollo llevaron a cabo el Taller sobre “Sistema de Indicadores para el Seguimiento de Conferencias Internacionales” (27 noviembre al 5 diciembre, 2002 en CEPAL-Santiago). El objetivo general del taller fue promover la adopción de un conjunto común de áreas temáticas e indicadores, de modo de aunar criterios y favorecer la comparabilidad internacional. Para ello se diseñaron dos módulos de trabajo, uno de propuesta de indicadores y otro de desarrollo informático. En el primero se discutieron los aspectos conceptuales, metodológicos y operacionales relativos a los indicadores, mientras que en el segundo se capacitó al personal técnico de los Institutos Nacionales de Estadística en el uso de Redatam para obtener indicadores, como en *R+xPlan* para la programación de un Sistema Nacional de Indicadores, en base a la propuesta Regional, elaborando un prototipo del sistema para cada país. A este taller, se invitó a los diez países de América del Sur y está previsto repetir esta actividad en mayo del 2003, invitando al resto de los países de América Latina. ✠

Figura 3.5. Al ingresar al sistema mediante la pantalla de figura 3.1, se accede en las pantallas para El Cairo (superior) y Beijing (inferior) a perfiles de países, etc.



Versión revisada en inglés de un artículo en español de la revista *Notas de Población*, No.75 (Santiago, Chile, CELADE/CEPAL, diciembre 2002). Se puede encontrar el artículo en español en: <http://www.eclac.cl/celade/>.

## The REDATAM Software for Census Data Dissemination and Analysis<sup>1</sup>

Dirk Jaspers Fajier ([djaspers@eclac.cl](mailto:djaspers@eclac.cl))  
Serge Poulard ([spoulard@eclac.cl](mailto:spoulard@eclac.cl))  
CELADE-ECLAC, Santiago, Chile

The REDATAM software system consists of a family of products which have been developed by CELADE to provide effective tools both for disseminating census, survey and other numerical data and for their processing by a wide range of users, such as National Statistical Office (NSO) analysts, national and regional planners, local authorities, academic researchers, etc.

This paper presents some of the main characteristics of the current fourth generation Redatam family, *Redatam+G4* (abbreviated *R+G4*)<sup>2</sup>. Particular emphasis is given to the different options that system offers to process and analyze census data and the corresponding channels that can be used to reach groups of users with differing needs and access to information technology.

### What does *R+G4* do?

The *R+G4* capabilities are built around a standard kernel: the *R+G4 statistical engine*. This set of highly efficient routines is used to create or import databases into a proprietary format—one of the security features of the system—to produce required outputs and to generate new variables that can be aggregated permanently to the database. External databases are generally converted to the proprietary format that compresses, encrypts and inverts the original data in order to combine efficiency and confidentiality of information, i.e., rather than a single file with records of the values of each variable for each “questionnaire”, there is a separate file for each variable containing the value of the variable on each “questionnaire”.

Originally designed to facilitate local planning by allowing efficient tabulation of population census data for user-defined small *ad-hoc* areas, the scope of Redatam has been widened to permit processing other types of data such as agricultural censuses, household surveys, etc. The software continues to be particularly efficient for defining and processing data for small areas as required by local-level planners and analysts in the public and private sectors.

To take full advantage of census and other data, a *R+G4* database is usually constructed with *microdata*, that is, the variables refer to individuals, households, or other elements of analysis, so that the variables can be combined in tabulations for any geographical

area defined by the user. The data are organized hierarchically making it possible to derive, for example, new household variables based on the number and characteristics of the individuals in the house. Any set of sub-areas can be selected in such a way that only the selected subset of data is processed, thus speeding up the overall processing to produce various types of output tabulations. Furthermore, any variables that are derived may be permanently saved into the database for future use. Although there is a powerful easily-learned Redatam programming language, no knowledge of programming is required to get quick results which are selected in graphical windows.

The program contains facilities to process and display maps connected with the geographical database levels. This allows the user to display spatial analysis of the aggregated statistics that are produced by the system for any of the geographical levels.

*R+G4* runs under Microsoft Windows 95, 98, 2000, NT4 and XP. The previous version, *winR+ v1.2* (and *ZonPlan*), can also be used under Windows 3.x.

### What are the members of the Redatam family?

The Redatam software family developed by CELADE, presently has the following members available<sup>3</sup>:



***R+Process*** for processing data from user-selected areas, including the entire database, with programs written in the Redatam command language, or using the more limited Assistants—similar to Wizards in some commercial software—that do not require any programming skills nor any knowledge about the Redatam language. This module also includes facilities for expanding, appending, connecting, downloading and other management of existing *R+G4* databases.



***R+Create*** for creating proprietary format *R+G4* hierarchical databases from files in ASCII, xBase, IMPS, ISSA, or CHILLAN (internal format of *R+G4*).



***R+xPlan*** for the construction, by programmers, of applications for distribution to end-users that permit them to obtain indicators, often involving complex calculations, without their needing any knowledge of Redatam or the Redatam language. When naming an application for a specific Redatam database, the *x* may be replaced by a descriptive name; e.g., *R+AgroPlan* (or *AgroPlan* for short).



***R+WebServer***, the latest (2002) member of the family, for requesting information from an *R+G4* census database on an intranet, extranet or on the Internet.

The first three are shipped together in the *R+G4* package. It and the *R+Webserver* can be downloaded, free of charge, from the Celade website: <http://>

## What are the main characteristics of the R+G4 software and databases?

As noted above, one of the main objectives behind the original development of Redatam in the early 1980's was to meet the need of the Latin American NSO's for processing specific *ad hoc* small-area tabulations that could not be included with the official census results. This demand, identified in a needs study of institutions in selected countries, led CELADE to design a database management tool that would manage large volumes of census microdata with hierarchical (geographical) structure down to the smallest area of the census administrative exercise, often city blocks or similar-sized areas.

The first generation of Redatam, which appeared in 1986, could process a single census; this restriction was removed in the next version to allow the combination of information from several censuses and/or other data sources. Additionally, the targeted users were extended beyond the NSOs and now comprise a variety of other national institutions, including ministries and regional and local governmental agencies, as well as, universities, research institutions and the private sector. By the early 1990's, copies of the 2<sup>nd</sup> generation of the software were sent —most were sold—, to persons and institutions in around 80 countries. With respect to the present 4<sup>th</sup> generation, approximately 8250 downloads of the system and its updates have been made since its release on the Internet in August 2001 to the present (November 2002).

The following sub-sections briefly outline the main characteristics of the fourth generation of *R+G4*:

### **Statistical outputs from existing and derived variables**

You can obtain frequencies, cross tabulations, averages, counts and AreaLists (results for a list of geographical areas) from a Redatam database. The variables may be those which are already in the database, or which are derived, via recoding, combination or other manipulation from existing variables or which are in a linked external database. If you need

more advanced statistics, you can use *R+G4* to extract the required information from the database and then export the data to an analysis software (see below).

### **Hierarchically structured databases and associated dictionaries**

A Redatam database is characterized by its hierarchical structure. In the case of a census database, the structure generally is based on the administrative-political organization of the country with additional census area(s) at the lowest level to facilitate the enumeration, for example:

Country  
Region  
Province  
Municipality  
District  
Enumeration Area

Below the lowest of these geographical levels, Enumeration Area in the example, the hierarchy continues downward, reflecting the census forms, usually:

Dwelling  
Household  
Person

The database dictionary contains information on these levels (known as “entities”) along with the description on the variables for each of the entities. The names of each of the areas, known as “elements”, making up an entity (e.g., the names of each of the provinces, or in the case of the areas without names, their codes) can also be made available to users (see “Selection of geographical areas...” below).

### **Multi-source multi-census databases**

A Redatam database can be created by joining several databases from different sources, leading to a multi-sectoral hierarchical database. The demonstration database from the hypothetical country New Miranda provided with the software has such a multi-branch

Visitenos en nuestro sitio web: <http://www.eclac.cl/celade>

structure as seen in Fig. 4.1. Information from two censuses is combined with information on education (at the block level) and on agriculture (at the comuna level), along with the relevant metadata for each of these.

It is also possible to link the database with external data sources without physically integrating the information in a single Redatam database. For example, simple instructions can link an existing database to an external xBase file with compatible geographical codes.

### Data compressed to 25% of the original

Compression of the hierarchically structured data files and the use of an internal Redatam format makes the size of a Redatam census database much smaller than the size of the original data in ASCII format (normally around 25% of the original).

### Confidentiality protection for databases

While the NSO, as owner of the database, has full access to the geographical, housing, household and person levels and to the entire set of variables, the access of other users can be restricted according to confidentiality regulations of the specific NSO, for example,

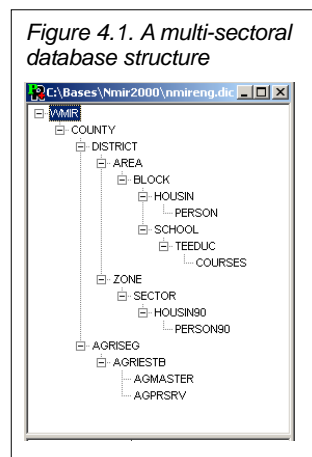


Figure 4.1. A multi-sectoral database structure

end-users might only be able to see results at the District or Municipality level. To this end, the Redatam software includes a “Wizard” to facilitate the downloading of (sub) census databases according to specific criteria that define the lowest geographical level that can be accessed, the geographical area covered, the selection of variables, etc. This feature provides

a “safe” means to distribute census data to regional and local authorities and others. This feature has made it possible for NSOs to disseminate their microdata to third parties.

### Very high processing velocity

The hierarchical structure together with the internal data storage method and optimization results in extremely fast data processing, making *R+G4* highly efficient for whole country processing; a hierarchical tabulation of sex by type of housing for an entire Venezuelan census of 23 million persons and 6 million dwellings on a 950Mhz computer took 82 seconds, or around 420,000 records per second. Since the system reads and processes only the data for user-selected areas, it is particularly efficient for working with specific geographical areas and for producing results for sub-areas within a given area (an Arealist) for, say,

mapping the spatial distribution of a variable.

### Selection of geographical areas to process

Taking advantage of the hierarchical structure of levels (“entities”) in a Redatam database, the system provides lists of names and codes of all selectable areas with facilities so that you can quickly define a “selection” of *ad-hoc* geographical areas. Areas can also be defined from digital maps and according to quantitative criteria calculated from variables in the database. A selection can be composed of any number and mixture of the areas utilized in the census. The values for one or more variables (known as an Arealist) can be obtained for each and every sub-area of a selection, for example, the number of children under 5 years in each enumeration area within a municipality.

### Creation of new variables using the hierarchical database structure

In addition deriving new variables from existing variables at a given level, e.g., recoding or adding the values of variables, the hierarchical structure allows you to mix data at different levels (such as combining variables at the dwelling level with those at person level to create new variables at the dwelling level) or to aggregate data to higher levels (such as counting the number of single-parent families in each district of a province).

### Powerful & easy programming language

The programming language of *R+G4* is a straightforward easily learned object-oriented language, which has only three basic instructions (RUNDEF, DEFINE and TABLE), each of which can have various sub-clauses. Additionally, “visual programming” is offered since each basic instruction may be edited with an Assistant.

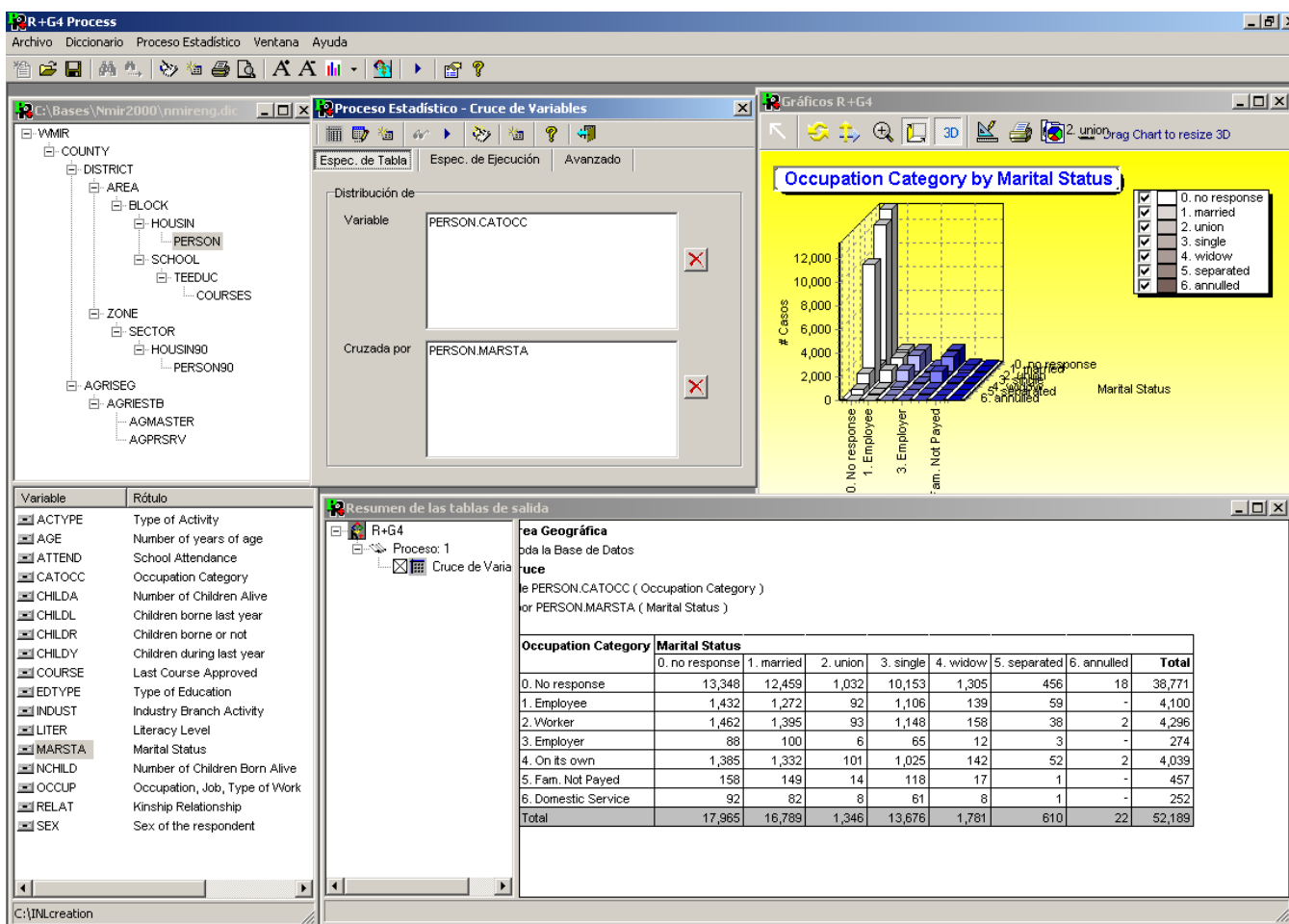
### Facilities for users with different proficiency levels

Although knowledgeable users can take advantage of the *R+G4* programming language, you can obtain results without writing any instructions at all. Frequencies, cross tabulations, averages, counts and arealists (results for each of a list of geographical areas) can be obtained by novices with “EasyTabs” (Assistants or “Wizards”), which only require dragging and dropping existing variables of interest from the dictionary window to the “Easy” window (see Figure 4.2). Even expert users often use the EasyTabs to obtain results quickly that do not need the manipulations and processing possible with the programming language. It is also possible to use the *R+xPlan* module of the software to create applications adapted to the needs of end-users who have no knowledge of Redatam (see the section on *R+xPlan* below).

### Graphics, mapping and export facilities

In addition to a host of features too long to list here, *R+G4* allows you to produce a variety of graphs and, to display the results of AreaLists on thematic maps to better visualize spatial relationships. That is, you

Figure 4.2. A table and related graphic obtained using Easy Tabs for tabulations. Selecting an entity (left-upper window) shows its variables (left-lower window) which are dragged & dropped in the middle-upper window defining the tabulation.



can prepare a list of results of a given variable for any sub-areas with a study zone, such as the percentage of houses with cars in, say, each district or city block within a municipality, and then classify the results into a set of categories, assign a color to each, display the results on a digital map and label and organize the map for printing.

If you wish to make more sophisticated analyses of the results produced by R+G4, you can export the output to specialized software like Microsoft Excel for working with spreadsheets, the Statistical Package for Social Science (SPSS) for advanced statistical analysis, and Geographical Information Systems (GIS) for spatial analysis.

As mentioned previously, it is very easy to create tailored sub-databases of a given database to facilitate data dissemination and detailed analysis in other institutions, for example, providing cities or municipalities with their own census data. Naturally, all the necessary metadata in the dictionary are included, so the sub-database from the point of view of its user is a standard Redatam database.

### On-line Help Manuals in three languages

There is on-line Help in English, Spanish and Portuguese (and manuals in these languages in .pdf format can be downloaded for printing). The internal structure of the software permits easy translation into other languages; external institutions took advantage of this in previous Redatam generations to produce French and Vietnamese translations.

### What are countries doing with their Redatam census databases?

Since the late 1980's Redatam has been extensively used in the region and elsewhere for processing census microdata, both to request results from whole censuses, taking advantage of the high compression of the data and the high velocity of processing, and to request information on local and other geographical sub-areas, taking advantage of the area selection and the capability to read only the data in the selected areas as well as the AreaList facilities that allow spatial display on maps within Redatam and export to external GIS.

Reasons such as these have led most Latin American and Caribbean NSOs to create Redatam databases of their 1980 and 1990-round censuses and most expect

Figure 4.3. Example of an *R+xPlan* application for the Costa Rica 2000 Population and Housing Census. Map of dependencia for districts.

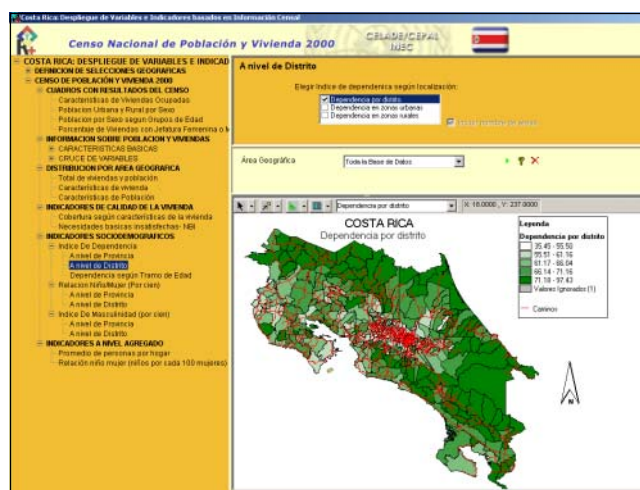
that Redatam will be one of the avenues of dissemination for their 2000 round census. Various African and Asian countries are also using Redatam for census data analysis and dissemination. The NSO use census (and other Redatam) databases for analysis, attention to visitors who request specific tabulations, for preparing specific sub-databases and for providing databases or sub-databases to third parties. For example, Cambodia sells the whole 1998 census database in Redatam format, while Chile provided each of its over 330 municipalities with their own 1992 census data in Redatam format.

### What is a *R+xPlan* application?

The *R+xPlan* module, as noted earlier, offers developers the means to create applications that can be tailored to the needs of different end-users for data dissemination and publishing. It is not necessary for the user of an *R+xPlan* application to understand how Redatam works or the internal characteristics of a database to utilize the database; query parameters are selected from a user-friendly and straightforward interface. In addition to presenting results in tabular format, graphs and thematic maps can be produced automatically, with the possibility to edit the results or export them to specialized external software.

Unlike working directly in the Redatam software, access to the data and dictionary in an *R+xPlan* application is pre-defined by the developer and new variables and ad-hoc tabulations cannot be derived by the user. On the other hand, an application allows the end-user, without programming, to take advantage of complex indicators provided by the developer (such as those relating to family structure, unmet basic needs, etc.). The institution owning the data, and usually creating the *R+xPlan* application, can include restrictions on what can be done with the data while customizing the application to meet the needs of specific end-users, without providing extensive, or in most cases, any training. On-line help, together with metadata, can be incorporated in the application to facilitate its use.

Specialized statistical or demographic analysis techniques can also be implemented in an application (although for the time being this can only be done by CELADE programmers working with the source code). This allows applications to be developed without exporting the data to third-party specialized software. For example, in-house applications have been created to calculate income distribution and poverty indicators (such as Gini, Atkinson, Theil and Sen indices) from census and household surveys microdata (see, for example, Ortiz y Faúndez, 2001) "Indicadores de Desigualdad y Pobreza y su Cálculo en Redatam", *Redatam Informa*, vol 7, December 2001, pp. 9-14). Another inhouse example is the indirect estimation of infant and child mortality and fertility, from census



and survey data based on the information on children ever born, surviving children and children born last year. Unseen by final users, the *R+xPlan* engine receives its instructions from simple text files, known as .INL files, that control the whole application (see Carrasco, 2001). Consequently, the creation of an *R+xPlan* application is quite easy, and can be taught in a few days to professionals who know how to create and process Redatam databases.

### Examples of *R+xPlan* applications

The following are some of the *R+xPlan* applications which have been developed by various institutions for distribution on CD's or which are currently under development:

#### Microdata

##### Population and census data

- Costa Rica: censuses of 1973, 1984 and 2000 (see Fig. 4.3)
- Panama: 2000
- Mongolia: 2000

##### Agricultural censuses

- Brazil: 1996 (for internal use)
- Chile: 1997 (for internal use)

##### Industrial Survey/census

- Chile: 1995-1997 annual surveys (economic variables are declared as sensitive and no individual data from enterprises can be obtained);

##### Vital Statistics

- Chile: 1997-1999 (births and deaths)

#### Aggregated data

##### International trade statistics (internal use)

- ComerPlan (world trade statistics since 1970)

##### Follow-up indicators

- Programme of Action, International Conference on Population and Development (ICPD), Cairo, 1994: Latin America and the Caribbean pre- and post-Cairo indicators.



<http://www.eclac.cl/id.asp?id=10865> for download; see p. 18 of this issue for the Internet version URL. See pp.10 -11 on creation of national indicator systems for summit conference followup.

### Indicators for social planning and monitoring

- SID Panama (all institutions in the government's Social Action Programme contribute; also includes the UNDP Human Development Indicators). [www.gabsoc.gob.pa](http://www.gabsoc.gob.pa).

### What are the characteristics of a R+WebServer application?

The *R+WebServer* software, provided free of cost, triggered high interest after its appearance as a beta version in mid-2001, since it is an inexpensive way for an NSO to "disseminate" census data—and eventually other data—in comparison with the cost of purchasing and implementing commercial software. "Dissemination" in this case uses an extranet/intranet or the Internet to allow remote processing of data stored on a server in a centralized databank without giving internal or external users access to the original data files.

Another advantage for processing databases on-line within a NSO is that the processing speed does not depend on the PC from which the query is initiated. The response time depends only on the network and the characteristics of the server which are generally adequate in most NSO.

The access to, and options for, data processing on-line via an intranet or the Internet are totally controlled by the database owner, and the nature of the access and data options (including a possible nominal fee) may be adapted to the types persons using the system. Moreover, while providing databases on CDs to third parties implies the risk of illegal copying (and loss of possible income), the on-line data processing provides much more security.

The currently available *R+WebServer* is designed to run in the Windows environment—in the near future there will also be a version for the UNIX/LINUX platform—. Like an *R+xPlan* application, it uses a standard Redatam database and is implemented through one or more ASCII files that describe the management of the site. The site can provide access to different databases and the same database can be accessed by

several Web applications according to the needs to be met. For example, an application, using census data, could focus on physical planning while another one could focus on socio-demographic themes only.

In practice, the Web site developer creates a set of request pages from which the user selects a predefined set of parameters such as the variables to be tabulated, geographical area of interest, type of output, etc. The user-selected parameters are sent to the server that converts these parameters into a query (statistical process), executes the request, formats the output and returns the resulting page to the user's browser. Data processing options are still somewhat limited (although complex R+ programs can be sent to and processed on the server); if resources become available, a client-server application will be developed, which will open a new range of options for processing census and related data at-a-distance.

Two to three days of instruction on how to create *R+WebServer* applications is sufficient for technical persons who have Redatam+ data processing experience.

### Examples of R+WebServer applications

#### Intranet

##### Databanks

- ECLAC intranet: The CELADE Regional Census Databank which includes all the censuses available in Redatam format. Several household surveys, agricultural censuses and other databases can be also be accessed.
- ECLAC Sub-regional Headquarters for the Caribbean is creating a databank for the Caribbean region.
- Venezuela: The NSO has a databank with the censuses of 1971, 1981 and 1990, recent household and fertility surveys and vital statistics.

##### Individual censuses

- The NSO of Brazil, Costa Rica, Panama, Mongolia, Saint Lucia, Saint Vincent, Grenada and others have at least one data base available on-line via their local intranets.



### Internet Censuses

- Saint Lucia: 1991 census tabulations (<http://www.stats.gov.lc/>): Two applications are offered, one for making *ad-hoc* tabulations, the other for standard tabulations agreed upon by the Caribbean countries for the 2000-round of censuses, but allowing users to specialize any given tabulation, such as by filtering and recoding variables.
- Costa Rica: 1973, 1984 and 2000 censuses (<http://www.inec.go.cr/>. Mirror site: <http://www.eclac.cl/censos>, then: click “Procesar datos en-linea”.
- Curacao, Grenada, St. Vincent and other Caribbean countries: censuses.

### Indicators

- Panama: SID. <http://www.eclac.cl/celade/censos>, then click “Procesar datos en linea”. (See also the *R+xPlan* application on p. 16).
- CELADE: Follow-up indicators for the Programme of Action of the International Conference on Population and Development (ICPD), Cairo, 1994 (cf. the *R+xPlan* application above).

### Summary and conclusions

The Redatam system family provides three options for disseminating census data: 1) The **R+Process** module of the *R+G4* software, the data dictionary and the census database together with all relevant census documentation can be placed on a compact disk (CD) to allow users full data access, at a given level of protection, through the Redatam software and its programming language; 2) The **R+xPlan** module of the *R+G4* software can be used to create applications that can be placed on a CD to provide a simple way for end-users to obtain pre-defined indicators with some user specification for any geographical areas from the census and other data without knowing how to use the Redatam programming language; 3) The **R+WebServer**, via the Internet or an intranet, can provide end-users with pre-defined indicators similar to those available from *R+xPlan* applications on CD, but with much greater protection for the data since only the results are provided to the user.

As emphasized in this article, the desired level of access can always be set in a given application by each NSO or database owner. An additional advantage of the *R+xPlan* and the *R+WebServer* options is that they allow the design of tailored applications for specific users and can more easily be written in the local language — as has been done by the NSO of Mongolia — since they have a much smaller number of screens than the entire *R+G4* software, itself.

At present, Internet data processing should not be the only option used to disseminate of census results, since the Internet is not yet universal, transmission speeds are

often low and costs high. Therefore, it is recommended that a mixture of alternatives provided by the *Redatam+G4* family be used to disseminate census data.

Current and future development of the Redatam family is aimed at improving the interface, with facilities in the software for further demographic and statistical analysis, as well for applying demographic analysis techniques. Data processing capabilities of the *R+WebServer* are being amplified (with filters, recoding schemes, geographical selections, etc.) and, in the future, a Redatam client-server will be offered to allow more ample utilization of the *R+G4* capabilities on the Internet. Finally, the Redatam family will also operate on the UNIX/LINUX platform.

### Bibliography

- Carrasco, Sebastián, 2001 “Construcción de una aplicación *R+xPlan*: Un ejemplo basado en los censos 1974-2000 de Costa Rica”, *Redatam Informa*, vol. 7, diciembre 2001, pp. 2-6. <http://www.eclac.cl/id.asp?id=9736>
- CELADE “2000 Censuses” sub-site: <http://www.eclac.cl/>
- CELADE home page: <http://www.eclac.cl/celade>
- CELADE “Redatam Informa” sub-site: <http://www.eclac.cl/celade/redatam>
- Ortiz, Laura y Faúndez, Juan Eduardo, 2001. “Indicadores de Desigualdad y Pobreza y su Cálculo en Redatam”, *Redatam Informa*, vol. 7, diciembre 2001, pp. 9-14. <http://www.eclac.cl/id.asp?id=9736>

### Notas

<sup>1</sup> Revision of a paper presented in English at the 20th Asian Population Census Conference, Ulaanbaatar, Mongolia, 19-21 June, 2002.

<sup>2</sup> Celade began the development of Redatam in the mid-1980’s with a grant from the International Development Research Centre IDRC) of Canada, which provided equally generous grants for creating the second and third generations of the system (released, respectively, in 1991 and 1997; see also note 3). Support during this work was also obtained from projects financed by the Canadian International Development Agency, the United Nations Population Fund (UNFPA), the Interamerican Development Bank (IDB) and the Regular Budget of the United Nations. Major support for the present fourth generation and continuing effort comes from the Regular Budget of the United Nations with additional financing from the Regional CELADE-UNFPA and IDB (TC-00-0205-4-RG) projects.

<sup>3</sup> In addition to these Redatam members, as part of the joint CELADE-University of Waterloo (Ontario, Canada) project during the mid-1990’s that was financed by IDRC, the University created a Redatam-related group of spatial decision support tools using an in-house Geographical Information System (GIS). Information on these tools, *AccessPlan* for primary health and family planning, *EduPlan* for educational planning, and *TourPlan* for for tourism-environment planning, can be obtained from [www.fes.uwaterloo.ca/Tools/](http://www.fes.uwaterloo.ca/Tools/) or [gball@kupe.uwaterloo.ca](mailto:gball@kupe.uwaterloo.ca). ☒



Población y Salud Reproductiva

[www.DevelopmentGateway.org](http://www.DevelopmentGateway.org)



**Valor:** Todo el software Redatam es **gratis** y usted puede hacer las copias que desee.

**Idiomas:** El usuario puede adaptar el programa al español, inglés o portugués.

**DESCARGANDO ("DOWNLOAD"):** Esta es la forma **preferida** de entrega, pues es la más rápida y no tiene costo. Siga las instrucciones de: [www.eclac.cl/celade/redatam](http://www.eclac.cl/celade/redatam).

**Requerimientos mínimos para el programa:** Computador compatible con IBM; Pentium 133 mhz o superior; 32mb RAM; monitor a color SVGA; disco duro con al menos 17 mb para el programa y el ejemplo de base de datos, más el espacio para la base de datos propia; lector de CD-ROM; impresora (a color para mapas) Windows 95, 98, NT4, 2000 ó XP (*winR+* v1.2 trabaja *también* con Windows 3.1).

**Envío por correo:** El programa en CD es gratis, con un cargo de US\$25 por CD, por gastos del CELADE. Llene el formulario y mándelo al CELADE, cheque a la orden de ECLAC girado sobre un banco de los EE.UU. (En Chile, pague en pesos chilenos al cambio de las NU). Envío por courier: agregue US\$30 por cada orden de Latinoamérica y el Caribe y US\$50 para el resto del mundo.

**Más información:** Véase: [www.eclac.cl/celade/redatam](http://www.eclac.cl/celade/redatam).

Para obtener el software gratis por descarga (*download*), véase la sección superior.  
Para obtener el software en CD, por favor, indique la cantidad de CD en la primera columna.

Cantidad de CD	Programa	Descripción	Windows
	<b>Redatam+G4</b> ® <i>R+Process</i> <i>R+ Create</i> <i>R+ xPlan</i>	<b>Los módulos de la cuarta generación de Redatam:</b>  Procesa datos organizados jerárquicamente para áreas seleccionadas del usuario, mediante el Asistente "señale y haga clic"; tiene capacidad total de programación con el lenguaje de comandos Redatam+ que permite definir nuevas variables y resultados complejos. Mapea resultados temáticos. Exporta resultados en varios formatos y a SIG. Permite también que el usuario pueda expandir, anexas, desagregar y empalmar bases de datos <i>R+G4</i> .  Crea bases de datos jerárquicas desde archivos ASCII, xBase, IMPS, ISSA, o CHILLAN (formato de <i>R+G4</i> ).  Hace posible crear aplicaciones para el usuario final con indicadores calculados para una base de datos <i>R+G4</i> determinada; los usuarios de una aplicación <i>R+xPlan</i> no necesitan conocer el lenguaje ni el programa Redatam.	95 ó más
	<b>R+G4 Web Server</b> ®	Procesa bases de datos con Redatam a través de Internet; requiere un servidor web de Windows.	95 ó más
	<b>winR+</b> ® con <b>ZonPlan</b>	Versión previa; 3ra generación. Se recomienda utilizarla solamente para trabajar con computadores con Windows 3.x.	3.1, 95 ó más
	<b>Programas de apoyo winR+GIS para decisiones espaciales</b> <sup>1</sup>	<b>AccessPlan</b> ®: acceso a atención primaria de salud y planificación familiar; <b>EduPlan</b> ®: planificación educacional; <b>TourPlan</b> ®: Planificación de desarrollo turístico. Todos están diseñados para usar resultados de software Redatam+.	3.1, 95 ó más

<sup>1</sup> Solicítelo directamente a: Faculty of Environmental Studies, University of Waterloo, Ontario, Canada N2L 3G1.

Contacto: [gbhall@kupe.uwaterloo.ca](mailto:gbhall@kupe.uwaterloo.ca). Véase también: [www.fes.uwaterloo.ca/Tools/](http://www.fes.uwaterloo.ca/Tools/)

RI2002\_SolicitudSoftw7e.cue 3/03/03 16:31

Para obtener el software gratis por descarga (*download*), véase la sección superior.

Para obtener el software en CD:

Adjunto cheque en dólares estadounidenses a la orden de ECLAC, por la suma de:

\_\_\_ ejs. del CD @ US\$25 c/u = US\$ \_\_\_ + gastos envío, si hay (vea información arriba). Total = US\$ \_\_\_\_\_

Nombre: \_\_\_\_\_ Cargo: \_\_\_\_\_

E-mail: \_\_\_\_\_ División/Institución: \_\_\_\_\_

Dirección postal: \_\_\_\_\_ Ciudad: \_\_\_\_\_ País: \_\_\_\_\_

**Envíelo al CELADE a la casilla postal indicada a continuación.**



NACIONES UNIDAS



**Centro Latinoamericano y  
Caribeño de Demografía  
(CELADE / CEPAL)**

Casilla 179-D, Santiago, Chile

Fax: (56-2) 208-0196

Tel: (56-2) 210-2000

email: [Redatam@eclac.cl](mailto:Redatam@eclac.cl)