

NOTAS DE POBLACIÓN

AÑO XXIX, N° 75, SANTIAGO DE CHILE



NACIONES UNIDAS

C E P A L

Comisión Económica para América Latina y el Caribe
Centro Latinoamericano y Caribeño de Demografía (CELADE) – División de Población

LC/G. 2186-P
Diciembre de 2002

Copyright © Naciones Unidas 2003
Todos los derechos están reservados
Impreso en Chile

La autorización para reproducir total o parcialmente esta obra debe solicitarse al Secretario de la Junta de Publicaciones. Sede de las Naciones Unidas, N.Y.10017, EE.UU. Los Estados miembros y sus instituciones gubernamentales pueden reproducir esta obra sin autorización previa. Sólo se les solicita que mencionen la fuente e informen a las Naciones Unidas de tal reproducción.

PUBLICACIÓN DE LAS NACIONES UNIDAS

NÚMERO DE VENTA: S.03.II.G.77

ISBN 92-1-322063-4
ISSN impreso 0303-1829
ISSN electrónico 1681-0333

Ilustración de portada: Curdo Bermúdez, "El balcón" (detalle), 1941.
Gentileza del Banco Interamericano de Desarrollo
Diseño de portada: María Eugenia Urzúa

COMISIÓN ECONÓMICA PARA AMÉRICA LATINA Y EL CARIBE
José Antonio Ocampo Secretario Ejecutivo

CENTRO LATINOAMERICANO Y CARIBEÑO DE DEMOGRAFIA
(CELADE) – DIVISIÓN DE POBLACIÓN
Miguel Villa, Oficial a cargo

La Revista **NOTAS DE POBLACIÓN** es una publicación del Centro Latinoamericano y Caribeño de Demografía (CELADE) - División de Población, cuyo propósito principal es la difusión de investigaciones y estudios de población sobre América Latina y el Caribe, aun cuando recibe con particular interés artículos de especialistas de fuera de la región y, en algunos casos, contribuciones que se refieren a otras regiones del mundo. Se publica dos veces al año (junio y diciembre), con una orientación interdisciplinaria, por lo que acoge tanto artículos sobre demografía propiamente tal, como otros que aborden las relaciones entre las tendencias demográficas y los fenómenos económicos, sociales y biológicos.

Comité editorial:

Jorge Bravo
Juan Chackiel
José Miguel Guzmán
Rolando Sánchez
Susana Schkolnik

Coordinador Técnico:

Juan Enrique Pemjean

Secretaria:

María Teresa Donoso

Redacción y administración:

Casilla 179-D, Santiago, Chile
E-mail: mdonoso@eclac.cl

Precio del ejemplar: US\$ 12

Suscripción anual: US\$ 20

Las opiniones expresadas en esta revista son responsabilidad de los autores, sin que el CELADE sea necesariamente participe de ellas.

SUMARIO

Presentación	7
Foreword	10
Avant-propos	13
América Latina: los censos del 2000 y el desarrollo social. <i>José Miguel Guzmán y Susana Schkolnik</i>	17
Los censos en América Latina: nuevos enfoques. <i>Juan Chackiel</i>	45
Cómo adaptar las nuevas tecnologías a las operaciones censales. <i>Arij Dekker</i>	73
Algunas reflexiones acerca de un formulario censal diseñado para la captura de datos mediante tecnología de imágenes (escáner). <i>Nelly T. Niedworok</i>	115
Los censos de población en Internet. <i>Dr. Werner Haug</i>	133
El censo de Canadá de 2006. Dirección estratégica. <i>Anil Arora</i> ..	147
¿Podrán los censos de los países en desarrollo responder a las futuras demandas de información?: el caso de Argentina. <i>Alejandro Giusti</i>	165
Preguntas para investigar la fecundidad y la mortalidad en la niñez y la mortalidad de adultos. <i>Ralph Hakkert</i>	195
La población indígena en Costa Rica según el censo del año 2000. <i>Elizabeth Solano Salazar</i>	217

¿Vale la pena hacer encuestas de post-empadronamiento censal? <i>David C. Whitford y Jeremiah P. Banda</i>	259
Preservación de archivos con documentos y microdatos censales y aumento de los grupos de gestión. <i>Wendy L. Thomas y Robert McCaa</i>	303
El software Redatam para divulgación y análisis de datos censales. <i>Dirk Jaspers Faijer y Serge Poulard</i>	321

EL SOFTWARE REDATAM PARA DIVULGACIÓN Y ANÁLISIS DE DATOS CENSALES

(Versión revisada del documento presentado en la 20ª Conferencia sobre
Censos de Población, Ulán Bator, Mongolia,
19 al 21 de junio de 2001)

Dirk Jaspers Faijer* y **Serge Poulard****
CELADE - CEPAL

RESUMEN

REDATAM es un sistema de software compuesto por un conjunto de módulos desarrollados por el Centro Latinoamericano y Caribeño de Demografía (CELADE) - División de Población de la CEPAL, cuyo objetivo consiste en ofrecer herramientas efectivas para divulgar datos censales, de encuestas y otros datos numéricos, y que estos puedan ser analizados por una amplia variedad de usuarios, como analistas de las oficinas nacionales de estadísticas, oficiales de planificación nacional y regional, autoridades locales e investigadores académicos. El software es fácil de usar y permite analizar, visualizar y publicar datos con más rapidez y con mayor eficiencia de costos que cualquier otro programa de su categoría. Puede adaptarse a usuarios con distintos niveles de capacitación, ya que ofrece rutinas sencillas, en las que bastan unas selecciones con el ratón para definir una tabla, así como un potente lenguaje de programación con el que se pueden generar indicadores sofisticados y realizar complejos tabulados y manipulaciones de datos. En el software también destacan la

* djaspers@eclac.cl.

** spoulard@eclac.cl.

velocidad de procesamiento y la posibilidad de seleccionar áreas definidas por el usuario (incluso las áreas de enumeración del censo), lo cual permite obtener, con el mínimo costo, el máximo provecho de censos y otras bases de datos de gran tamaño. A su vez, se ha incorporado de forma específica la protección de la confidencialidad para apoyar la divulgación de la actual ronda de censos. Estas funciones se están poniendo a disposición de todas las oficinas nacionales de censos de América Latina y el Caribe y de otras instituciones de ámbito nacional.

En este documento se presentan algunas de las características más importantes de la cuarta generación de la familia Redatam, *Redatam+G4* (o bien *R+G4* en su forma abreviada).¹ Se hace hincapié en las diferentes opciones que ofrece el sistema para procesar y analizar datos censales y los canales correspondientes que pueden usarse para llegar a grupos de usuarios con necesidades dispares y distintos niveles de acceso a la tecnología informática.

¹ El CELADE - División de Población de la CEPAL, inició el desarrollo de REDATAM a mediados de la década de los ochenta con una donación del Centro Internacional de Investigación para el Desarrollo (CIID) de Canadá, que también financió proyectos igualmente generosos para crear la segunda y la tercera generación del sistema (presentadas en 1991 y 1997, respectivamente; véase también la nota 3). Durante esta labor también se contó con apoyo proveniente de proyectos financiados por la Agencia Canadiense de Desarrollo Internacional (ACDI), el Fondo de Población de las Naciones Unidas (UNFPA), el Banco Interamericano de Desarrollo (BID) y el presupuesto ordinario de las Naciones Unidas. El apoyo principal para la actual cuarta generación y para la continuidad del proyecto proviene del presupuesto ordinario de las Naciones Unidas con financiamiento adicional de proyectos regionales del CELADE con el UNFPA y el BID (TC-00-02-05-4-RG).

ABSTRACT

REDATAM is a software system composed of a family of modules which has been developed by CELADE to provide effective tools 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. The easy-to-use software allows users to analyze, visualize and publish data faster and more cost-efficient than other software in its category. It can accommodate users with different proficiency levels, by offering “Easy” routines that require little more than mouse clicks to define a table, as well as a powerful programming language that can be used to generate sophisticated indicators and carry out complex data manipulations and tabulations. Emphasis has also been placed both on processing speed and the selection of user-defined areas down to census enumeration areas so that very large census and other databases can be economically and fully exploited, while protecting confidentiality have been incorporated specifically to support the dissemination of the present round of census and are being made available to Latin American and Caribbean National Statistical Offices (NSOs) and other national institutions.

This paper presents some of the main characteristics of the current fourth generation REDATAM family, *Redatam+G4* (abbreviated *R+G4*).¹ 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.

¹ CELADE began the development of REDATAM in the mid-1980s 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 (CIDA), the United Nations Population Fund (UNFPA), the Inter-American 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.

RÉSUMÉ

REDATAM est un logiciel composé d'un ensemble de produits mis au point par le Centre latino-américain et des Caraïbes de démographie (CELADE) – Division de la population de la CEPALC visant à offrir des outils efficaces de divulgation de l'information censitaire, d'enquêtes et d'autres données numériques, de façon à ce que cette information puisse être analysée par une vaste gamme d'utilisateurs tels que les analystes des bureaux nationaux de statistiques, les fonctionnaires de la planification nationale et régionale, les autorités locales et les chercheurs universitaires. Le logiciel est d'un usage facile et permet d'analyser, de visualiser et de publier des données de façon plus rapide et moins coûteuse que tout autre programme similaire. Il peut s'adapter à des utilisateurs possédant différents niveaux de formation puisqu'il offre certaines routines simples à exécuter pour lesquelles le marquage par la souris suffit à définir un tableau, ainsi qu'un langage puissant de programmation qui permet de produire des indicateurs sophistiqués et réaliser des tabulations et des manipulations complexes de données. Ce logiciel se caractérise également par la vitesse de traitement et la possibilité, pour l'utilisateur, de sélectionner certains domaines spécifiques (notamment d'énumération du recensement), ce qui permet, à un coût minimum, de tirer le meilleur parti possible des recensements et d'autres bases de données de grande envergure. D'autre part, des dispositifs spécifiques ont été introduits pour assurer la confidentialité durant les phases de publication et de divulgation de la série actuelle de recensements. Ces fonctions sont mises à la disposition de tous les bureaux nationaux de recensements d'Amérique latine et des Caraïbes, ainsi que d'autres institutions à l'échelon national.

Ce document contient une présentation de certaines caractéristiques principales de la quatrième génération de Redatam, *Redatam+G4* (ou *R+G4* en abrégé).¹ L'accent est mis sur les différentes options qu'offre le système

¹ Le CELADE – Division de la population de la CEPALC, a commencé la mise au point du Redatam au milieu des années grâce à une bourse du Centre de recherches pour le développement international du Canada qui a également octroyé des bourses tout aussi généreuses pour la création des deuxième et troisième générations du système (présentées respectivement en 1991 et 1997, voir aussi la note 3). Ces travaux ont également été soutenus par des projets financés par l'Agence canadienne de développement international (ACDI), le Fonds des Nations Unies pour la population (FNUAP), la Banque interaméricaine de développement (BID) et le budget ordinaire des Nations Unies. La principale source de soutien pour cette quatrième génération et la continuité du projet est le budget ordinaire des Nations Unies et le financement additionnel de projets régionaux du CELADE-FNUAP et de la BID (TC-00-0205-4-RG).

pour traiter et analyser les données censitaires et les voies pouvant être utilisées pour parvenir à des groupes d'utilisateurs qui présentent des besoins divergents et différents niveaux d'accès à la technologie informatique.

I. ¿Qué hace *R+G4*?

Las capacidades de *R+G4* están construidas sobre la base de un núcleo estándar denominado “motor estadístico *R+G4*”. Este conjunto de rutinas de alta eficiencia se utiliza para crear o importar bases de datos en un formato propio —esta es uno de los elementos de seguridad del sistema— y para obtener los productos requeridos y generar variables nuevas que pueden agregarse de forma permanente a la base de datos. Habitualmente, las bases de datos externas se convierten al formato propio de Redatam, que comprime, encripta e invierte los datos originales con el fin de combinar la eficiencia con la confidencialidad de la información.

El diseño inicial del programa obedecía en parte a la necesidad de que la vida útil de los datos censales de vivienda y población abarcara el período de diez años que media entre una ronda censal y la siguiente; sin embargo, el ámbito de Redatam se ha ampliado para dar paso al procesamiento de otros tipos de datos, como los censos agrícolas y las encuestas de vivienda. El motor estadístico permite procesar millones de registros en una computadora personal en cuestión de minutos, o en unos segundos si se utilizan microcomputadoras más potentes. El software es particularmente eficiente en el procesamiento de información de zonas pequeñas, por lo que se adapta a las necesidades habituales de los encargados de planificación y los analistas de ámbito local, tanto del sector público como del privado.

Para obtener el máximo provecho de los datos censales o de otra naturaleza, las bases de datos *R+G4* se suelen construir con microdatos, es decir, variables que hacen referencia a individuos, hogares u otros elementos de análisis; estas variables pueden combinarse después en tabulados para mostrar resultados para cualquier área geográfica que defina el usuario. Los datos se organizan jerárquicamente, lo cual permite derivar, por ejemplo, nuevas variables de hogar a partir del número y las características de los individuos que habitan en cada casa. Se puede seleccionar cualquier conjunto de subáreas para que sólo se procese el subconjunto elegido de datos, lo cual acelera el cálculo global necesario para generar varios tipos de tabulados. Además, cualquier variable derivada puede guardarse de forma permanente en la base de datos para usarla en el futuro. No se requiere ninguna experiencia previa de programación para obtener tabulados rápidos y otros resultados estadísticos, ya que la interfaz funciona mediante ventanas gráficas.

El programa contiene recursos para procesar y crear mapas conectados con los niveles geográficos de la base de datos. Esto permite que el usuario vea en la pantalla un análisis espacial de las estadísticas agregadas que el

sistema produce en cualquiera de los niveles geográficos. Las funciones de visualización, junto con la manipulación de datos y la alta velocidad de procesamiento, facilitan el acceso rápido a los datos y aumentan su valor.

El software *R+G4* funciona en sistemas operativos Windows 95, 98, 2000, NT4 y XP. La versión anterior, *winR+ v1.2* (y *ZonPlan*) puede utilizarse también con Windows 3.x.²

II. ¿Quiénes componen la familia REDATAM?

La familia de software REDATAM,³ desarrollada por el CELADE, está compuesta en la actualidad por los módulos siguientes:

R+Process, para procesar datos de áreas seleccionadas por el usuario o de toda la base de datos, mediante potentes programas escritos en el lenguaje de comandos de Redatam, o bien mediante los Asistentes —también conocidos como «Easies»—, más limitados, que no requieren conocimientos de programación ni nociones del lenguaje de Redatam. Este módulo también contiene opciones para expandir, añadir, conectar y descargar bases de datos de *R+G4*, así como otras funciones de gestión.

R+Create, para crear bases de datos jerárquicas en el formato *R+G4* a partir de archivos ASCII, xBase, IMPS, ISSA o CHILLAN (formato interno de *R+G4*). El usuario de este módulo es el propietario de los datos originales.

R+xPlan, para que los programadores puedan construir aplicaciones de distribución que permitan a los usuarios finales obtener indicadores, que puedan requerir cálculos complejos, sin necesidad de

² *winR+ v1.2* es la tercera generación de REDATAM. Dado que *R+G4* sólo puede importar un diccionario de *winR+ v1.2* (archivo .mdb), los usuarios que quieran trabajar en *R+G4* con bases de datos de primera o segunda generación, es decir, con diccionarios .def, .con y .lbl generados por Redatam Plus para DOS o Redatam v3.1, tendrán que convertir antes los diccionarios al formato .mdb con *winR+ v-12*. Quienes trabajen con Windows 3.x (entorno de 16 bits) también deberán usar antes *winR+ v1.2*.

³ En el proyecto conjunto CELADE-Universidad de Waterloo (Ontario, Canadá), desarrollado a mediados de la década de los noventa y financiado por el Centro Internacional de Investigación para el Desarrollo (CIID), la Universidad creó un grupo de herramientas Redatam para apoyar la toma de decisiones (Spatial Decision Support Tools) utilizando un sistema de información geográfica (SIG) propio. Se puede obtener más información sobre estas herramientas (*AccessPlan* para salud primaria y planificación familiar, *EduPlan* para planificación educativa y *TourPlan* para planificación turística-ambiental) en el sitio web <http://www.fes.uwaterloo.ca/Tools> o escribiendo a la dirección gbhall@kupe.uwaterloo.ca.

conocimientos sobre Redatam o sobre su lenguaje. Al dar nombre a una aplicación para una base de datos específica, la *x* puede cambiarse por un nombre descriptivo; por ejemplo, una aplicación basada en un censo agropecuario podría llamarse *R+AgroPlan* (o sencillamente *AgroPlan*).

Estos tres componentes de la familia Redatam vienen en un paquete conjunto como módulos del software *R+G4* que puede descargarse gratuitamente desde el sitio web del CELADE: <http://www.eclac.cl/celade/redatam>.

El servidor web Redatam+

R+WebServer, el último de los miembros de la familia (aparecido en 2002), permite realizar consultas a una base de datos *R+G4* a través de una red intranet o extranet, o bien por Internet. Como sucede con los demás módulos, el software *R+WebServer* está basado en el motor estadístico Redatam, que desarrolla una alta eficiencia en entornos de red.

III. ¿Cuáles son las características principales del software R+G4 y sus bases de datos?

Uno de los objetivos principales que motivó el desarrollo del software Redatam a principios de la década de los ochenta fue la necesidad que tenían las oficinas nacionales de estadísticas latinoamericanas de procesar tabulados específicos de áreas pequeñas que no podían incluirse en los resultados oficiales de los censos. Esta demanda, identificada durante un estudio sobre las necesidades de las instituciones de varios países, llevó al CELADE a diseñar una herramienta de gestión de bases de datos que fuera capaz de manipular grandes cantidades de microdatos censales con una estructura jerárquica (geográfica) que permitiera descender hasta el área geográfica más pequeña definida en el ejercicio administrativo del censo, que a menudo es una manzana o un área de un tamaño equivalente. Además, el software debería funcionar en una computadora personal, procesar los datos con rapidez, requerir el mínimo espacio de disco posible y ser fácil de usar.

La primera generación de Redatam, que apareció en 1986, sólo podía procesar un único censo; esta restricción fue eliminada en la versión siguiente, en la que se podía combinar información de varios censos o de otras fuentes de información. Además, las oficinas nacionales de estadísticas

dejaron de ser los únicos usuarios interesados al añadirse una amplia variedad de instituciones nacionales, tales como ministerios, organismos gubernamentales locales y regionales, universidades, institutos de investigación y empresas del sector privado. A principios de la década de los noventa se enviaron copias de la segunda generación del software —la mayoría se vendieron— a personas e instituciones de unos 80 países; la cuarta generación y sus actualizaciones han sido descargadas gratuitamente por cerca de 10 000 personas (desde su presentación en el sitio web del CELADE en agosto de 2001 hasta el momento de escribir estas líneas, a fines de 2002).

Las secciones siguientes describen brevemente las características principales de la cuarta generación del sistema:

Productos estadísticos a partir de variables existentes y derivadas

Se pueden obtener frecuencias, tabulados cruzados, medias, conteos y listas de área (resultados que se presentan para cada uno de los elementos de una lista de áreas geográficas) a partir de una base de datos de Redatam. Las variables pueden ser las que ya existen en la base de datos o las que se hayan derivado de ella mediante reprogramación, combinación o manipulación de variables existentes, o de otras que contenga una base de datos externa enlazada. Si se necesitan productos estadísticos más avanzados, se puede usar *R+G4* para extraer la información requerida de la base de datos y exportar los datos a un formato legible por el software de análisis (véase más abajo).

Bases de datos y diccionarios estructurados jerárquicamente

Una base de datos de Redatam se caracteriza por su estructura jerárquica. En el caso de una base de datos censal, la estructura suele responder a la organización político-administrativa del país con áreas censales adicionales en el nivel más bajo para facilitar la enumeración, como en el siguiente ejemplo:

País
 Región
 Provincia
 Municipio
 Distrito
 Área de enumeración

Por debajo de los niveles geográficos inferiores, es decir, del «Área de enumeración» que aparece en el ejemplo, la jerarquía continúa descendiendo y refleja los propios cuestionarios del censo, que suelen estar estructurados en:

Vivienda
Hogar
Persona

El diccionario de la base de datos contiene información sobre estos niveles (denominados «entidades») junto con la descripción de las variables de cada una de las entidades. Los nombres de cada una de las áreas (denominadas «elementos») que forman una entidad (por ejemplo, los nombres de cada una de las provincias o, en el caso de áreas sin nombre, los códigos) también pueden ponerse a disposición de los usuarios (véase más adelante la sección «Selección de áreas específicas»).

Bases de datos de origen múltiple y/o de varios censos

Una base de datos de Redatam se puede crear uniendo varias bases de datos procedentes de fuentes diferentes, lo cual generaría una base de datos jerárquica multisectorial. Este tipo de estructura se ilustra mediante la base de datos de demostración **Nueva Miranda** (nombre de un país imaginario; este archivo se suministra con el software). Se combina la información de dos censos con datos de educación (al nivel de manzana) y agricultura (al nivel de comunas). Además, el software permite incluir metadatos relevantes para la(s) base(s) de datos.

También es posible enlazar la base de datos con fuentes de datos externas sin necesidad de integrar físicamente la información en una única base de datos de Redatam. Por ejemplo, con instrucciones muy simples se puede enlazar una base de datos con un archivo de formato xBase, siempre y cuando la identificación geográfica de cada registro sea compatible con la que se definió en la base de datos de Redatam.

Compresión de datos al 25% del tamaño original

La compresión de los archivos de datos estructurados jerárquicamente y el uso de un formato de archivo propio hacen que el tamaño de una base de datos censal de Redatam sea mucho más reducido que el del original en formato ASCII (normalmente un 25% del original).

Protección de los datos confidenciales

La oficina nacional de censos, propietaria de la base de datos, tiene acceso total a los niveles geográficos, de vivienda, familiares y personales, y al conjunto completo de variables; sin embargo, el acceso de los demás usuarios puede restringirse en función de las normas de confidencialidad que dictamine la oficina en cuestión. Por ejemplo, los usuarios pueden tener autorización para ver sólo los resultados de los niveles Distrito y Municipio. Con este fin, el software Redatam incluye un «Asistente» que facilita la descarga de subbases de datos censales de acuerdo con los criterios específicos de acceso para el usuario, donde se define el nivel geográfico más bajo que puede consultar, el área geográfica disponible, su selección de variables, etc. Este es un método seguro para distribuir datos censales a las autoridades locales y regionales y a otras instituciones conforme a las normas de confidencialidad de las oficinas nacionales de censos, y ofrece al mismo tiempo los datos requeridos para la planificación sectorial y local. Esta función ha permitido a las oficinas nacionales de censos divulgar sus microdatos a otros organismos e instituciones.

Alta velocidad de procesamiento

La estructura jerárquica, junto con el sistema de almacenamiento interno de Redatam (en el que los datos asociados a cada variable conforman un archivo aparte), permite procesar los datos a muy alta velocidad (como ejemplo ilustrativo, una computadora con procesador Pentium III a 900 MHz puede procesar alrededor de un millón de registros por segundo para obtener una frecuencia, y de 500 000 registros por segundo para generar un tabulado), por lo que resulta factible usar *R+G4* para procesar la información de todo un país. Dado que el sistema sólo lee y procesa los datos del área que haya seleccionado el usuario, es especialmente eficiente para trabajar con zonas pequeñas y para generar información sobre subáreas comprendidas en un área dada (es decir, listas de área).

Selección de áreas geográficas específicas para su procesamiento

Gracias a la estructura jerárquica de los niveles («entidades») de las bases de datos de Redatam, el sistema suministra las listas de nombres y códigos de todas las áreas seleccionables y las herramientas para definir con rapidez una «selección» de áreas geográficas específicas; para mejorar

su eficiencia, el software lee y procesa sólo los datos contenidos en la selección. También se pueden definir áreas a partir de mapas digitales y mediante criterios cuantitativos calculados a partir de las variables que contenga la base de datos. Una selección puede estar compuesta por cualquier número y tipo de áreas. Se pueden obtener los valores de una o más variables para todas y cada una de las subáreas contenidas en la selección (denominados «listas de área»), por ejemplo, se podría consultar el número de niños de menos de cinco años en cada área de enumeración de un municipio determinado.

Creación de variables nuevas a partir de la estructura jerárquica de la base de datos

Además de derivar variables nuevas a partir de las que ya están definidas en un nivel dado (por ejemplo, para recodificar los valores de una variable), la estructura jerárquica permite mezclar datos en niveles diferentes (por ejemplo, combinar variables del nivel Vivienda con los del nivel Persona para crear variables nuevas en el primero de ellos) o agregar datos en niveles superiores (por ejemplo, contar el número de familias monoparentales que hay en cada distrito de una provincia dada).

Lenguaje de programación potente y fácil de aprender

El lenguaje de programación de *R+G4* es orientado a objetos y resulta muy sencillo, pues sólo tiene tres instrucciones básicas (RUNDEF, DEFINE y TABLE), cada una de las cuales puede ir acompañada de varios subcomandos. Además, se ofrece «programación visual», ya que cada instrucción básica puede editarse con un Asistente.

Funciones para usuarios con diferentes niveles de conocimiento

Aunque los usuarios más experimentados pueden aprovechar al máximo el lenguaje de programación de *R+G4*, es posible obtener resultados sin escribir ni una sola instrucción. Los principiantes pueden extraer frecuencias, tabulados cruzados, medias, conteos y listas de área (resultados para cada elemento de una lista de un área geográfica) utilizando los Asistentes (Easies), para los que sólo se requiere tomar las variables que se necesiten de la ventana del diccionario, y arrastrarlas hasta la ventana del Asistente. Los usuarios avanzados también recurren a menudo a este medio para obtener con rapidez los resultados para los que no se requiere

toda la capacidad de gestión y procesamiento que ofrece el lenguaje de programación.

También es posible usar el módulo *R+xPlan* para crear aplicaciones adaptadas a las necesidades de los usuarios finales que no tengan conocimientos de Redatam (véase el capítulo V).

Gráficos, mapas y opciones de exportación

Además de su conjunto de opciones, demasiado numerosas para enumerarlas aquí, *R+G4* permite generar una amplia variedad de gráficos y mostrar los resultados de las listas de área en mapas temáticos para representar mejor las relaciones espaciales. Esto significa que se puede preparar una lista de resultados de una variable dada para cualquier subárea de una zona (como por ejemplo el porcentaje de casas que tienen automóvil en cada distrito o manzana de un municipio), y a continuación clasificar los resultados conforme a un conjunto de categorías, asignar un color a cada una, mostrar los resultados en un mapa digital, etiquetarlo y organizarlo para imprimir.

Si se van a hacer análisis más sofisticados a partir de los resultados producidos por *R+G4*, se pueden exportar los productos resultantes a software especializado, como Microsoft Excel para trabajar en hojas de cálculo, SPSS para realizar análisis estadísticos avanzados, y sistemas de información geográfica (SIG) para llevar a cabo análisis espaciales.

Como se dijo más arriba, es muy fácil crear bases de datos a medida a partir de una base de datos principal para facilitar la divulgación de los datos y su análisis detallado en otras instituciones, por ejemplo para entregar a ciudades o municipios sus propios datos censales. Como es natural, todos los metadatos del diccionario estarán incluidos en la nueva base de datos, de forma que a efectos de los usuarios funcionará como una auténtica base de datos de Redatam.

Ayuda en línea y manuales en tres idiomas

La ayuda en línea y los manuales en formato PDF están disponibles en inglés, español y portugués. Los manuales pueden descargarse e imprimirse. La estructura interna del software permite una traducción sencilla a otros idiomas; algunas instituciones aprovecharon esta característica para traducir al francés y al vietnamita las versiones anteriores de Redatam.

IV. ¿Qué hacen los países con las bases de datos censales de Redatam?

Redatam se ha usado masivamente para el procesamiento de microdatos censales, tanto para obtener resultados de censos completos (aprovechando la alta tasa de compresión de los datos y la alta velocidad de procesamiento) como para obtener información de alcance local o de otras subáreas geográficas (gracias a las funciones de selección de área, a la capacidad para leer solamente los datos de las áreas seleccionadas, y a las funciones de lista de área que permiten exponer sobre mapas los resultados obtenidos con Redatam y exportarlos a SIG externos).

Estas ventajas han llevado a la mayoría de las oficinas nacionales de estadística latinoamericanas a crear bases de datos censales en formato Redatam desde la década de los ochenta hasta la actualidad; muchos de ellos prevén que Redatam será una de las vías principales de divulgación de datos para la presente ronda. Varios países de África y Asia están usando también Redatam para analizar y divulgar los datos de sus censos. Las oficinas nacionales usan bases de datos de censos (y otros tipos de bases de datos en formato Redatam) para realizar análisis, satisfacer a los visitantes que solicitan tabulados específicos, preparar bases de datos a medida y suministrar bases de datos a terceros. Por ejemplo, Camboya comercializa la base de datos con el censo completo de 1998 en formato Redatam, y Chile suministró a cada uno de sus 330 municipios sus propios datos censales de 1992 en este formato.

V. ¿Cuáles son las características principales de una aplicación *R+xPlan*?

Como se dijo más arriba, el módulo *R+xPlan* ofrece a los programadores la posibilidad de crear aplicaciones a la medida de los distintos usuarios que participarán en la divulgación de datos. No es necesario que el usuario de una aplicación *R+xPlan* entienda las características internas de la base de datos; los parámetros de consulta se seleccionan mediante una interfaz sencilla y de uso intuitivo. Además de presentar los resultados en formato tabular, se pueden generar automáticamente gráficos y mapas temáticos, con la posibilidad de modificar el resultado o exportarlo a un formato de software especializado.

Al contrario de lo que sucede en el software Redatam, en una aplicación *R+xPlan* el acceso a los datos y al diccionario está definida por

el programador, lo que impide que el usuario derive variables nuevas y tabulados específicos. Por otra parte, las aplicaciones permiten, sin necesidad de programar, aprovechar los indicadores complejos suministrados por el programador (por ejemplo, los que se refieren a la estructura familiar o a las necesidades básicas insatisfechas, entre otros). La institución que sea propietaria de los datos, y que por lo general será también la que cree la aplicación *R+xPlan*, puede restringir la manipulación de datos y personalizar el funcionamiento de la aplicación para que se adapte a las actividades específicas que vayan a realizar los usuarios, sin necesidad de un extenso curso de capacitación, o incluso sin ninguna en absoluto. La ayuda en línea, junto con los metadatos, se puede incorporar en la aplicación para facilitar su uso.

Aunque los usuarios finales no puedan verlo, el motor de *R+xPlan* recibe sus instrucciones de archivos de texto con extensión *.inl* que controlan toda la aplicación.⁴ En consecuencia, la creación de una de estas aplicaciones es bastante sencilla; los profesionales que ya sepan crear y procesar bases de datos Redatam pueden aprender este proceso en pocos días.

En la aplicación se pueden incluir también técnicas especializadas de análisis estadístico o demográfico (aunque por el momento esto sólo pueden hacerlo los programadores del CELADE que trabajan con el código fuente). Las aplicaciones se pueden diseñar sin necesidad de exportar los datos al formato propio de un software especializado. Por ejemplo, el CELADE ha creado aplicaciones propias para calcular indicadores de distribución del ingreso y de pobreza (tales como los índices de Gini, Atkinson, Theil y Sen) a partir de microdatos censales y de encuestas de vivienda.⁵ Otro ejemplo es la estimación indirecta de la mortalidad infantil y de la fecundidad a partir de datos censales y de encuestas, en base a la información sobre niños nacidos vivos, niños sobrevivientes y niños nacidos el último año.

Ejemplos de uso de las aplicaciones *R+xPlan*

A continuación se mencionan varias aplicaciones *R+xPlan* desarrolladas por diversas instituciones y distribuidas en CD, así como otras que se encuentran en desarrollo:

⁴ Véase Carrasco, 2001.

⁵ Véase, por ejemplo, Ortiz y Faúndez (2001), pp. 9-14.

Microdatos:

Datos de población y censo:

- * Costa Rica: censos de 1973, 1984 y 2000.
- * Panamá: 2000.
- * Mongolia: 2000.

Censos agrícolas (sólo para uso interno):

- * Brasil: 1996.
- * Chile: 1997.

Censos/encuestas del sector industrial:

- * Chile: encuestas anuales de 1995 a 1997 (las variables económicas se catalogaron como datos sensibles y no es posible obtener datos aislados de las empresas).

Estadísticas vitales:

- * Chile: 1997-1999 (nacimientos y defunciones).

Datos agregados:

Estadísticas comerciales internacionales (sólo para uso interno):

- * ComerPlan (con estadísticas de comercio de todo el mundo desde 1970).

Indicadores de seguimiento para el Programa de Acción de la Conferencia Internacional sobre la Población y el Desarrollo (CIPD), El Cairo, Egipto, 1994

- * América Latina y el Caribe (con indicadores antes de la CIPD y después de la CIPD). Se pueden descargar en <http://www.eclac.cl/id.asp?id=10865>.

Indicadores para la planificación y el monitoreo social (incluye indicadores para evaluar la aplicación de las recomendaciones internacionales adoptadas en las cumbres mundiales)

- * SID Panamá (todas las instituciones relacionadas con el Programa de Acción Social del gobierno contribuyen; este sistema

incluye también los indicadores de desarrollo humano del Programa de las Naciones Unidas para el Desarrollo, PNUD). <http://www.gabsoc.gob.pa>.

- * Nicaragua, Guatemala, Ecuador y otros países iniciaron aplicaciones *R+xPlan* similares, que serán un aporte importante para el mecanismo de revisión decenal de la CIPD.

VI. ¿Cuáles son las características de una aplicación *R+WebServer*?

La versión beta del software *R+WebServer*, que se suministrará gratuitamente, generó un gran interés al hacer aparición hacia finales del año 2001, dado que para las oficinas nacionales de censos constituye un método sin costo para divulgar los datos censales —y de otra naturaleza—, en contraste con el gasto que supone adquirir y adaptar software comercial. El programa usa una intranet, una extranet o la Internet para permitir el procesamiento a distancia de los datos almacenados en un servidor, en un banco de datos centralizado que impide el acceso de los usuarios (internos o externo) a los archivos de datos originales.

Una ventaja adicional para procesar bases de datos a través de una red en una oficina nacional de censos es que la velocidad de procesamiento no depende de la calidad de la computadora de la que parte la consulta. El tiempo de respuesta depende únicamente del estado de la red y de las características del servidor, que suelen ser las adecuadas en la mayoría de las instalaciones. Esto implica una reducción del tiempo de procesamiento y una disminución de la inversión en equipos.

El acceso a los datos y las opciones de procesamiento en red están totalmente controlados por el propietario de la base de datos, y la naturaleza del acceso y las opciones disponibles (lo cual incluye la posibilidad de solicitar el pago de una tasa) puede adaptarse a los tipos de usuarios del sistema. Además, y frente al riesgo de copia ilegal (y pérdida de ingresos potenciales) que implica la distribución de bases de datos en CD, el procesamiento en red ofrece mucha más seguridad.

La versión disponible de *R+WebServer* está diseñada para funcionar en el entorno Windows, y en el futuro próximo habrá también una versión para plataforma UNIX/LINUX. Al igual que las aplicaciones *R+xPlan*, este software utiliza una base de datos estándar de Redatam y se configura mediante uno o más archivos ASCII que describen la administración del

sitio. El programa puede dar acceso a bases de datos diferentes, y una misma base de datos puede estar disponible para consultas a través de varias aplicaciones web, según sean las necesidades del usuario. Por ejemplo, una aplicación que use datos censales podría orientarse a la planificación física, mientras que otra, con los mismos datos, podría dedicarse sólo a temas sociodemográficos.

En la práctica, el administrador del sitio web crea un conjunto de páginas de consulta a través de las cuales el usuario selecciona un conjunto predefinido de parámetros, que pueden ser las variables que quiere tabular, el área geográfica que le interesa, el tipo de producto que desea obtener, etc. Los parámetros que seleccione se enviarán al servidor, que los convertirá en una consulta (procesamiento estadístico), la ejecutará, formateará el producto resultante y enviará de vuelta la página con el resultado al programa navegador del usuario. Las opciones de manipulación de datos aún son limitadas (aunque se pueden enviar programas complejos de *R+* para que el servidor los procese); si en el futuro se contara con los recursos necesarios, se desarrollaría una aplicación cliente-servidor que abriría un nuevo abanico de posibilidades para el procesamiento a distancia de datos censales y otros similares.

Las personas que tengan experiencia en procesamiento de datos con *R+* sólo necesitarán de dos a tres días de instrucción para aprender a crear aplicaciones *R+ WebServer*.

Ejemplos de aplicaciones *R+ WebServer*

Intranet

Bancos de datos

- * Intranet de la CEPAL: Banco de Datos Regional de Censos del CELADE, que contiene todos los censos disponibles en formato Redatam. También se puede acceder a varias encuestas de hogares, censos agrícolas y otras bases de datos.
- * La sede subregional de la CEPAL para el Caribe está creando una base de datos para la región del Caribe.
- * Venezuela: la oficina nacional de censos tiene un banco de datos en el que están incluidos los censos de 1971, 1981 y 1990, las encuestas de hogares más recientes y las estadísticas vitales.

Censos

- * Las oficinas nacionales de censos de Barbados, Brasil, Costa Rica, Granada, Mongolia, Panamá, Santa Lucía y Trinidad y Tabago, entre otras, poseen al menos una base de datos en red disponible a través de sus intranet locales.

Internet

Censos

- * Santa Lucía: censo de 1991.
<http://www.stats.gov.lc/Scripts/RG4WebEngine.exe/PortalAction>
- * Costa Rica: censos de 1973, 1984 y 2000.
<http://www.inec.go.cr>
- * Barbados, Curaçao, Granada, Trinidad y Tabago y otros países y territorios caribeños: censos.

Indicadores

- * Panamá: SID (véase el apartado sobre las aplicaciones *R+xPlan*, más arriba).
- * CELADE: indicadores de seguimiento del Programa de acción de la Conferencia Internacional sobre la Población y el Desarrollo (CIPD), El Cairo, Egipto, 1994 (véase el apartado sobre las aplicaciones *R+xPlan*, más arriba).

VII. Resumen y notas finales

La familia de productos Redatam ofrece tres posibilidades principales para la divulgación de datos censales:

- * El **módulo *R+Process* del software *R+G4*** proporciona acceso total a cualquier base de datos de censo de Redatam y a sus diccionarios de datos; mediante el lenguaje de programación, los

usuarios pueden crear variables nuevas y, si lo desean, añadirlas como variables permanentes, y aprovechar todas las demás opciones de programación que ofrece Redatam. El propietario de los datos puede restringir el acceso de los distintos tipos de usuarios definiendo la unidad mínima de información geográfica, las variables y el ámbito geográfico que cada uno de ellos tendrá a su disposición. El diccionario de datos, la base de datos y el software pueden distribuirse en un CD junto con toda la documentación relevante del censo.

- * El módulo **R+Plan del software R+G4** se puede usar para crear aplicaciones que entrega una manera simple para usuarios finales para la obtención de indicadores predefinidos, sin conocer el lenguaje de programación de Redatam. Además, los resultados pueden ser graficados y presentados sobre mapas. El nivel de acceso a los datos es totalmente definido por el propietario de la base de datos. Las aplicaciones *R+xPlan* basadas en datos censales pueden variar de acuerdo a las características de los usuarios finales y pueden ser distribuidos en un CD junto con toda la documentación censal.

- * El programa **R+WebServer** es una forma aún más sencilla de divulgar datos censales y de otra naturaleza, y de proveer indicadores predefinidos para cualquier área geográfica a un usuario a través de una red. También se puede suministrar documentación censal a través del sitio web. El programa *R+WebServer* también permite crear gráficos y mapas temáticos, aunque, al contrario de lo que sucede con las aplicaciones *R+xPlan* en CD, el usuario no puede modificar estos productos. El propietario de la base de datos es quien define el nivel de acceso a los datos. Esta fórmula permite proporcionar resultados y no implica la entrega de los datos a terceros, por lo que no hay ningún riesgo de copia ilegal de la información. Puede definirse un nivel de acceso diferente para cada tipo de usuario; además, el acceso puede ser gratuito o de pago. Si se solicita a los usuarios que se registren, el propietario de los datos puede tener una idea de los usos que se están dando a sus datos y llevar a cabo las mejoras y adaptaciones necesarias.

Estas opciones plantean una disyuntiva entre el nivel de sofisticación de los cálculos posibles y la simplificación de la obtención de información. La capacidad analítica máxima se logra con *R+Process*, ya que da acceso al diccionario Redatam, permite derivar variables nuevas y hacer cualquier

cosa que esté al alcance del lenguaje de programación Redatam. Esta capacidad de análisis es algo inferior en *R+xPlan* y en *R+WebServer*, puesto que se limita a las opciones que el programador haya querido incluir.

Se ha subrayado en este artículo que las oficinas nacionales de censos o los propietarios de las bases de datos siempre tienen la posibilidad de determinar el nivel de acceso en una aplicación dada. Otra ventaja de las opciones *R+xPlan* y *R+WebServer* es que permiten diseñar aplicaciones a medida para usuarios específicos, y puede ser adaptado al idioma nacional, como hizo, por ejemplo, la oficina nacional de censos de Mongolia.

En la actualidad, el procesamiento de datos a través de Internet no debería ser la única opción para divulgar información censal, puesto que esta red aún no tiene alcance universal, la velocidad de transmisión suele resultar lenta y los costos son considerables. Por lo tanto, se recomienda utilizar una mezcla de las alternativas que ofrece *Redatam+G4* para este fin.

En cuanto al desarrollo actual y futuro de la familia de productos Redatam, se plantean los objetivos de mejorar la interfaz, añadir funciones al software para perfeccionar los análisis demográficos y estadísticos, y aplicar técnicas de análisis demográfico. Las opciones para procesar datos con el *R+WebServer* se está ampliando (con filtros, recodificación, selecciones geográficas, entre otros) y, en el futuro, se ofrecerá una aplicación cliente-servidor para fomentar la utilización extensiva de las funciones de *R+G4* a través de Internet. Por último, la familia de productos Redatam también funcionará en plataformas UNIX/LINUX.

BIBLIOGRAFÍA

- 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 de 2001, pp. 2-6, <http://www.eclac.cl/id.asp?id=9736>.
- CELADE (s/f), *Censos 2000*, <http://www.eclac.cl/id.asp?id=7784>. (s/f), sitio web, <http://www.eclac.cl/celade>.
- ____ (s/f), *Redatam Informa*, varios números, <http://www.eclac.cl/celade/redatam>.
- Ortiz, Laura y Juan Eduardo Faúndez (2001), «Indicadores de desigualdad y pobreza y su cálculo en Redatam», *Redatam Informa*. vol. 7, diciembre, pp. 9-14, <http://www.eclac.cl/id.asp?id=9736>.