



Red7 Create



Módulo de generación de bases de datos

MATERIAL DE CLASES 1

REDATAM© es una aplicación informática desarrollada por el Centro Latinoamericano y Caribeño de Demografía (CELADE), que es la División de Población de la Comisión Económica para América Latina y el Caribe, (CEPAL), Naciones Unidas.
www.cepal.org/celade/

Contents

I. Una Base de Datos Redatam	3
I.1 Conceptos básicos y definiciones	3
<i>ENTIDADES</i>	3
<i>ENTIDADES SUPERIORES</i>	4
<i>ENTIDADES INFERIORES</i>	4
<i>ELEMENTOS</i>	4
<i>ELEMENTOS IDENTIFICABLES O SELECCIONABLES</i>	4
Código geográfico.....	5
<i>ELEMENTOS NO IDENTIFICABLES Y/O NO SELECCIONABLES</i>	5
<i>RAMAS Y NIVELES</i>	6
<i>VARIABLES</i>	7
I.2 La base de datos REDATAM	7
<i>PROYECTO (PRJX)</i>	7
<i>DICCIONARIO DE LA BASE DE DATOS REDATAM (.DICX)</i>	7
<i>ARCHIVOS DE DATOS REDATAM (.RBF)</i>	8
<i>ARCHIVOS DE PUNTEROS O ÍNDICES (.PTR)</i>	9
<i>ALMACENAMIENTO DE LOS DATOS DE ENTRADA</i>	9
II. Creando una base REDATAM	13
II.1 Etapas preliminares	13
<i>CONSIDERACIONES BÁSICAS PARA LA GENERACIÓN DE LA BASE DE DATOS</i>	13
<i>PASOS PARA LA GENERACION</i>	13
II.2 REVISION DE LA CORRELACION GEOGRAFICA DE LA BASE DE DATOS	17
II.3 Ejemplo de carga con datos CSV	18
<i>CREACIÓN DE LA ESTRUCTURA DE LA BASE DE DATOS</i>	18
II.4 Ejemplo de carga con datos CSPro	22
<i>CREACIÓN DE LA ESTRUCTURA DE LA BASE DE DATOS</i>	23
<i>GENERACIÓN DE LA BASE DE DATOS</i>	27
II.5 REVISION DEL INFORME DE GENERACION.....	31
III. Revisión de la base de datos con Redtama7 Process	33
<i>CONSISTENCIA DEL CONTENIDO DE LA BASE DE DATOS</i>	33
<i>REVISAR LA BASE DE DATOS</i>	33
<i>PROCESOS DE FRECUENCIAS Y UN PLAN DE TABULADOS</i>	34
IV. Módulo Redatam7 Admin para modificar una base de datos	35
<i>CORRECCIONES DE VARIABLES</i>	35
<i>ANTES Y DESPUÉS</i>	35
V. Anexo: Plan de tabulados y programas	38
Ejemplos para evaluar la consistencia: Evaluación de EDAD del hijo mayor del hogar (respecto al jefe).....	39
Otros indicadores para evaluar consistencia.....	43

I. Una Base de Datos Redatam

La creación de una base de datos en formato REDATAM constituye la etapa final de un proceso que permitirá disponer de la información sobre población u otras unidades de estudio, organizada jerárquicamente, para facilitar el análisis de sus atributos a fin de realizar análisis y diagnósticos sociodemográficos que ayudarán a la implementación de acciones contenidas en políticas y planes de desarrollo local y nacional.

El proceso de creación constituye, en sentido estricto, una tarea sencilla compuesta de una serie de pasos que no requieren de ningún tipo de conocimiento informático avanzado. Sin embargo, antes de llegar al momento de la creación de la base de datos se han debido preparar los datos de origen según la estructura que adoptará finalmente. Aun cuando el proceso de preparación de los datos antes de llegar a crear la base en formato REDATAM implica tareas diferenciadas según la fuente y organización de los datos originales existe una serie de conceptos, propios de REDATAM, a los que aquellos han de ajustarse.

I.1 CONCEPTOS BÁSICOS Y DEFINICIONES

Para trabajar con una base de datos REDATAM, es necesario primero entender mejor su estructura. A continuación explicamos algunos conceptos y definiciones básicas que nos ayudaran en el camino.

ENTIDADES

Entidades son conjuntos de objetos lógicos organizados en forma jerárquica en la base de datos. Una entidad, por ejemplo, puede ser el conjunto de provincias o ciudades en un determinado país.

En la Figura 1, se presenta una base de datos simple, que contiene cuatro entidades organizadas jerárquicamente: PAIS, PROVINCIA, DISTRITO y VIVIENDA.

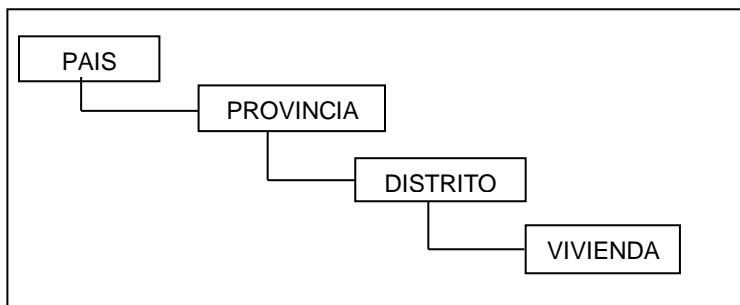


Figura 1. Estructura jerárquica simple con cuatro entidades

ENTIDADES SUPERIORES

Las entidades pueden ser clasificadas de acuerdo con su posición en la jerarquía. Las **entidades superiores** son aquellas que están sobre una entidad determinada. Tomando nuevamente la Figura 1 como referencia, podemos decir que la entidad DISTRITO es superior a la entidad VIVIENDA. La entidad inmediatamente superior puede ser llamada también de **entidad madre**. Por definición, cada entidad tiene sólo una entidad madre, y en consecuencia, sólo hay un único camino de vuelta hasta la entidad raíz.

ENTIDADES INFERIORES

Las **entidades inferiores**, a su vez, son aquellas entidades localizadas jerárquicamente debajo de una entidad determinada. En la Figura 1, se observa que la entidad DISTRITO es inferior a la entidad PROVINCIA. Las entidades inmediatamente inferiores pueden ser llamadas **entidades hijas**. Por definición, una entidad puede tener un número variable (igual o mayor que cero) de entidades hijas.

ELEMENTOS

Los **elementos** son los miembros individuales de cada entidad. Siguiendo con el ejemplo de la Figura 1, para la entidad PROVINCIA, los elementos son cada una de las provincias que componen el país.

El número de elementos varía significativamente de acuerdo a la entidad. Puede ser que existan unos pocos elementos, tal como las provincias, o varios millones como los individuos en la entidad PERSONA.

En el entorno estadístico, los elementos pueden asumir otros nombres, entre los más comunes pueden ser: observaciones, casos, instancias o registros. Para efectos didácticos, cuando nos referimos a una estructura de bases de datos, como la de REDATAM, optaremos por la expresión elemento. Para el caso de señalar el contenido de un archivo físico, usaremos la expresión registro.

Los elementos, así como las entidades, siempre respetan una jerarquía de estructura de bases de datos. Esto implica que cada elemento de una entidad madre está subdividido en los respectivos elementos de sus entidades hijas. Volviendo a la Figura 1, podemos deducir que cada elemento (provincia) de la entidad PROVINCIA está subdividido en elementos (distritos) de la entidad DISTRITO.

ELEMENTOS IDENTIFICABLES O SELECCIONABLES

Cuando los elementos de una entidad poseen nombres y/o códigos propios que los distinguen individualmente, decimos que son **identificables**. Al generar una base de datos en formato REDATAM, podemos elegir las entidades que contienen sus elementos identificados, razón por la cual son clasificadas como entidades **seleccionables**,

Aunque, por definición, cualquier entidad puede ser seleccionable, en la práctica, se acostumbra a asignar esta propiedad a las entidades geográficas debido a la posibilidad de analizar su distribución espacial. Por tratarse de entidades geográficas correspondiente a DAME División Administrativa Menor, desde el punto de vista de los formuladores de políticas públicas, existe interés en identificar sus elementos y obtener estadísticas para cada uno de ellos. En una base de datos cualquiera, por ejemplo, queremos saber cuál es el distrito más poblado de una provincia. Al marcar la entidad DISTRITO como seleccionable, REDATAM nos proporciona esta información relevante. Cada elemento en una estructura Redatam se puede identificar si se le asigna la característica de seleccionable, sin embargo, a niveles menores, se corre el riesgo de afectar el secreto estadístico. Cada INE al crear la base de datos Redatam, debe definir el nivel en donde no se afecte el secreto estadístico.¹

En el entorno de REDATAM, los elementos de las entidades seleccionables también pueden ser mapeados gráficamente usando su **código geográfico**.

Código geográfico

El código geográfico permite identificar en una tabla o mapa la unidad geográfica que dicho código esta identificando y el cual puede llegar hasta los niveles más desagregados de una base de datos ya sean las manzanas o sectores de enumeración. Los elementos personas, hogares y viviendas no llevan código geográfico por lo tanto no pueden ser identificados en una tabla o mapa).

El código geográfico está compuesto por una secuencia alfanumérica específica para cada uno de los elementos de una entidad. Cada parte del código informa la pertenencia a una determinada entidad, razón por la cual también se conoce como código compuesto. Un elemento del distrito 07, de la provincia 02, del país 0 (por definición), por ejemplo, puede recibir el código geográfico 00207.

ELEMENTOS NO IDENTIFICABLES Y/O NO SELECCIONABLES

Cuando los elementos de una entidad no tienen nombres y/o códigos propios que los distinguen individualmente, señalamos que son **no identificables**. Al generar una base de datos REDATAM, podemos elegir las entidades que no tendrán sus elementos identificables, razón por la cual son clasificadas como entidades **no seleccionables**. Este normalmente es el caso de entidades no geográficas, como vivienda y personas.

Aunque los elementos de entidades no geográficas no deben ser identificados individualmente, el procedimiento puede hacerse en conjunto. Esto se hace mediante la localización del grupo en elementos identificables de entidades superiores que son identificables. En el ejemplo anterior, un elemento específico probablemente una vivienda, que por definición, no es identificable, sin embargo, gracias a su código

¹ A esto se suma el hecho de que entidades menores -como personas-, en muchos países están protegidas por el principio del secreto estadístico que asegura la inviolabilidad de la intimidad de las personas físicas y jurídicas (empresas), las cuales no pueden ser objeto de divulgación de datos que las identifiquen estadísticamente.

geográfico – que comparte con otras viviendas del distrito- podemos identificarlo en su conjunto.

La característica de seleccionable para una entidad es definida por el usuario, al crear la entidad en el diccionario. También se usa el término de entidad Identificada y No Identificada, con el mismo propósito de Seleccionable y No Seleccionable, por la existencia de un código de identificación para sus elementos.

RAMAS Y NIVELES

Una **rama** es el camino recorrido desde la entidad raíz a una determinada entidad. Por definición, una entidad (y por ende todos sus elementos) pertenece a una única rama.

Una base de datos REDATAM puede tener más de una rama jerárquica y éstas a su vez pueden tener varias ramas, los conceptos y elementos de entidad son los mismos. Una entidad seleccionable no necesita ser necesariamente geográfica.

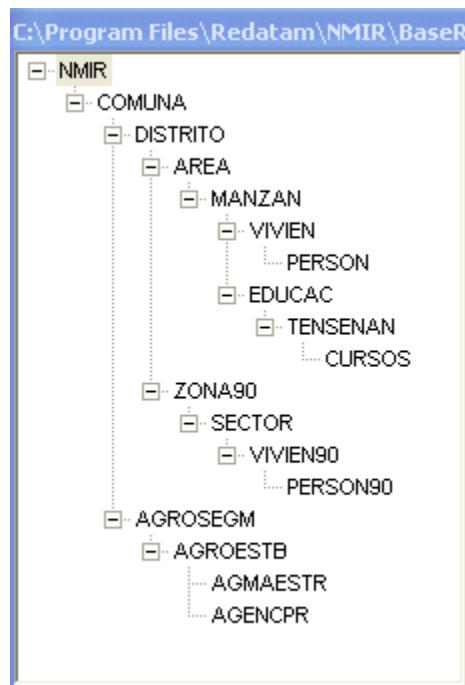


Figura 2. Estructura jerárquica multidisciplinaria

El **nivel** es el grado de profundidad de una entidad dentro de su rama. Por construcción, la entidad "raíz" tiene nivel cero, la(s) entidad(es) inmediatamente inferiores tienen nivel 1, las entidades bajo esta tienen el nivel 2 y así consecutivamente.

VARIABLES

Las variables son informaciones referentes a los elementos que integran una entidad y por ende una base de datos. Una base de datos censal guarda información en **variables**, tales como el sexo, parentesco y edad de las personas, el tipo de techo y pared de las viviendas, cantidad de milímetros de agua caída en las comunas, etc. Cualquier entidad de una base de datos puede contener variables, por ejemplo, la PROVINCIA puede tener variables como su nombre, código geográfico, total de personas, total de hogares, índice de vulnerabilidad, etc. El DISTRITO puede almacenar el número de viviendas sin agua potable o el total de jefes de hogar femenino.

La tabla 1, a continuación, muestra la relación entre los elementos y las variables de una entidad genérica.

ENTIDAD			
	Variable 1	Variable 2	Variable 3
Elemento 1 →			
Elemento 2 →			
Elemento 3 →			
Elemento 4 →			

Tabla 1. Relación entre elementos y variables de una entidad.

I.2 LA BASE DE DATOS REDATAM

Una base de datos REDATAM contiene necesariamente cuatro tipos de archivos, cada uno con características específicas: archivo del proyecto (.pejX), archivo del diccionario de datos (.dicX), archivos de datos (.rbf), y archivos de punteros (.ptr).

Los archivos de datos guardan las variables ya en formato REDATAM los cuales fueron cargados a través del módulo Red7 Create junto con los archivos de punteros.

PROYECTO (PRJX)

Se incorpora el proyecto como elemento administrador de la base de datos, diccionario, programas, selecciones geográficas, estilos de tablas, mapas y otros elementos. A través del proyecto se definen las carpetas donde se guardan cada uno de estos elementos.

DICCIONARIO DE LA BASE DE DATOS REDATAM (.DICX)

Este es un solo archivo que se genera para almacenar toda la metadata de la base de datos REDATAM, es decir la información de los datos. En el módulo Red7 Process se accede a la base de datos solamente a través del diccionario el cual incluye la estructura jerárquica, las entidades tanto seleccionables y no seleccionables, las claves de identificación, y las variables.

ARCHIVOS DE DATOS REDATAM (.RBF)

REDATAM posee funciones para leer y convertir datos guardados y escritos originalmente en formato CPro², SPSS³, xBase⁴, CSV⁵ y ASCII⁶ al formato REDATAM. Este proceso lee los metadatos desde su diccionario original y en base a la información que se define en un esquema REDATAM (.wipX), se ordenan los datos a la forma interna de REDATAM bajo una estructura binaria y comprimida con la extensión RBF (del inglés *Redatam Binary File*).

REDATAM guarda la información de cada variable en su propio archivo de datos, con un "registro" (línea) por cada elemento de la entidad. De ahí que cada variable es un vector de los datos. Existe un archivo para sexo, uno para estado civil, otro para tipo de vivienda, etc.

En la estructura de archivos de REDATAM el número total de registros (líneas) corresponde necesariamente al número de elementos (observaciones) que hay en la entidad a la que pertenece esa variable. En otras palabras, todos los elementos de una entidad deben tener un valor para esa variable en cuestión, ya sea un valor en blanco o ignorado. Para ilustrar esto se tiene la siguiente tabla con entidades y registros en cada una:

ENTIDAD	ELEMENTOS
PROVINCIA	5
DISTRITO	23
VIVIENDA	4.587

Tabla 2. Ejemplo de número de elementos por entidad.

Analizando la Tabla vemos que cada uno de los archivos de datos (archivo de variables) para la entidad PROVINCIA tiene 5 registros (líneas), para la entidad DISTRITO tienen 23 registros (líneas) y para la entidad VIVIENDA tienen 4.587 registros (líneas).

Los nombres que se asignan a cada uno de los archivos con variables están compuestos por el nombre de la base (definido al momento de creación de la base, por ejemplo, CPV2002) y una secuencia numérica ascendente. De esta forma, podemos tener los archivos CPV20020032.rbf para la variable sexo y CPV20020033.rbf para a variable edad.

² CPro es un programa estadístico, de dominio público, desarrollado por la Oficina de Censos de Estados Unidos (Census Bureau), la consultora Internacional ICF y SERPRO SA.

³ SPSS es un programa estadístico, de naturaleza comercial, desarrollado por la empresa homónima y comprador por la IBM en 2009.

⁴ xBase es el nombre genérico que se le da a las estructuras y lenguajes desarrollados a partir de dBASE, lenguaje comercial muy popular en la década de los 80.

⁵ CVS es el nombre que se le da a los archivos de datos que utilizan texto simple (comma separated values).

⁶ ASCII es un tipo de código binario usado mundialmente. Codifica un conjunto de 128 símbolos: 95 gráficos (letras del alfabeto, puntuación y símbolos matemáticos) y 33 símbolos de control.

ARCHIVOS DE PUNTEROS O ÍNDICES (.PTR)

Estos archivos son los responsables por la conexión entre los elementos de las entidades y sus entidades inferiores. Cada entidad tiene un archivo de punteros o índices, con elementos que "apuntan" desde los elementos de la entidad superior hasta los elementos de la entidad misma.

PERTENENCIA

El concepto de pertenencia es fundamental para entender el funcionamiento de los archivos de datos. En una estructura jerárquica, la pertenencia indica la relación de subordinación entre las entidades y sus elementos. La figura 3, muestra una estructura jerárquica genérica con cuatro entidades: PAIS, PROVINCIA, DISTRITO, VIVIENDA

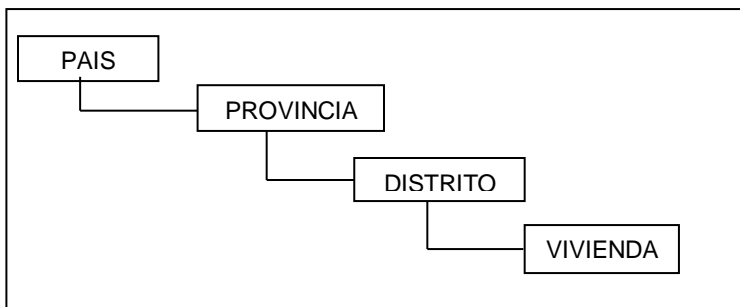


Figura 3. Estructura jerárquica

Analizando la figura 3 podemos decir que la entidad VIVIENDA "pertenece" a DISTRITO, que a su vez "pertenece" a PROVINCIA, que "pertenece" a PAIS, y esta, por definición, "pertenece" al sistema.

El concepto de "**pertenencia**", por ejemplo para una división político-administrativa, puede ser entendido como: la provincia 3 "del" país, el distrito 14 "de" la provincia 2, la vivienda 358 "del" distrito 21. Cada entidad puede tener un número distinto de elementos (casos), es decir, x distritos, y provincias, etc.

Para efecto de ejemplo, en la figura 3 se tiene PAIS como la entidad 0 (entidad raíz), PROVINCIA entidad 1, DISTRITO la entidad 2 y VIVIENDA la entidad 3.

Es importante mencionar que cada entidad puede tener diferente cantidad de elementos. Si retomamos la tabla 2, se observan los siguientes números de casos para las entidades: PROVINCIA tiene 5, DISTRITO tiene 234 y VIVIENDA tiene 4,587 elementos (numerados del 1 al 4,587).

ALMACENAMIENTO DE LOS DATOS DE ENTRADA

Para migrar una base de datos hacia REDATAM estos datos deben estar almacenados en uno de estos formatos CPro (cdf), SPSS (sav), dBASE (dbf), CVS y ASCII (txt) y deben estar relacionados a un diccionario de datos. Ahora bien, una vez realizada la migración a

REDATAM estos datos son convertidos y almacenados como archivos de datos binarios propios de REDATAM (rbf) junto con un diccionario (dicX) y los punteros (ptr).

Por otra parte los registros en un archivo externo, se pueden almacenar de dos maneras: con un formato único (SPSS, dBASE, CVS) o con un formato múltiple (CSPRO):

FORMATO ÚNICO O SIMPLE

El archivo de entrada tiene **un sólo tipo de registro**, esto es, todos los registros tienen el mismo formato e identifican a una misma entidad. En este caso, REDATAM asume que ellos pertenecen solamente a una entidad, por ejemplo PERSONA. Es posible mezclar variables de diferentes entidades en un archivo con un formato único siempre y cuando la entidad inferior este presente. Por ejemplo, si el archivo con formato único representa a personas, pero además se incluyen variables de la vivienda, todas las variables de la vivienda (luz, agua, piso, techo) deberán duplicarse en cada registro correspondiente a las personas dentro de esa vivienda (ver Tabla 3).

Identificación				Variables				
PROV	COM	DIST	VIV	PER	LUZ	AGUA	SEXO	EDAD
01	01	001	01	1	1	1	2	23
01	01	001	01	2	1	1	1	33
01	01	001	01	3	1	1	2	12
01	01	001	02	1	2	3	1	56
01	01	001	02	2	2	3	2	52
01	01	001	02	3	2	3	2	16

Tabla 3. Archivo de entrada de registros únicos (rectangular)

Para efectos de simplificación, a las entidades de índole geográfica (PROV, COM, DIST) solo se les ha asignado el código identificador. Este código es utilizado para mantener la **pertenencia** de una persona a una vivienda a un distrito a una comuna y a una provincia.

En el ejemplo se muestran las primeras personas pertenecen a la vivienda 01 y las otras tres personas siguientes pertenecen a la vivienda 02. Ambas viviendas pertenecen al distrito 001 que a su vez pertenece a la comuna 01 de la provincia 01. Por eso hay tanta repetición de códigos.

FORMATO MÚLTIPLE

Por otra parte, si se está tratando con un archivo de formato múltiple, tales como aquellos encontrados en encuestas de hogares, censos de población y agrícolas, **se tienen dos o más tipos de registros** en un mismo archivo. Por ejemplo, cuando se tiene un registro (una línea) para los datos de VIVIENDA y otro registro (una línea) para los datos de

PERSONA, este archivo debe tener una columna específica (la primera) para identificar el tipo de registro.

La tabla 4, la continuación muestra un ejemplo de un archivo de registros múltiples con dos tipos de registros señalados en la primera columna, "1" para vivienda y "2" para persona.

Tipo-reg	Identificación			Variables
	PROV			
	COM			
		DIST		
			VIV-PER	
	ID	ID	ID	ID.....
1	01	01	001	Vivienda 1
2	01	01	001	Persona 1
2	01	01	001	Persona 2
2	01	01	001	Persona 3
1	01	01	001	Vivienda 2
2	01	01	001	Persona 1

Tabla 4. Archivo de entrada de registros múltiples (tipo censos)

Nuevamente para efectos de simplificación, a las entidades de índole geográfica (PROV, COM, DIST) solo se les ha asignado el código identificador. Este código es utilizado para mantener la **pertenencia** de una persona a una vivienda a un distrito a una comuna y a una provincia.

En este caso, cada tipo de registro (el tipo 1 para vivienda, y el tipo 2 para personas) será referido a una entidad en la base de datos en formato REDATAM y los datos en cada tipo de registro serán cargados ordenadamente sólo a esa entidad. Cada tipo de registro y por ende la línea entera debe tener variables pertenecientes sólo a la entidad específica a que se refiere, más las variables de identificación (pertenencia) de las entidades de niveles superiores en la jerarquía como DISTRITO, COMUNA y PROVINCIA.

Para las entidades que no tienen variable de identificación, como es VIVIENDA y PERSONA, el sistema REDATAM "sabr " cu ndo crear un elemento detectando su tipo de registro en el archivo de entrada. Esto es, usando el mismo ejemplo de un t pico archivo de censos, cada tipo de registro "1" crear  un elemento vivienda para la entidad VIVIENDA, y los registros tipo "2" que vienen despu s de este registro tipo "1" crear  registros de personas para la entidad PERSONA manteniendo la pertenencia a la vivienda previa. Cuando encuentre otra vez un registro tipo "1" cerrar  la vivienda previa y abrir  una nueva vivienda para contener a las personas que pertenecen a esta nueva vivienda. Las viviendas sin personas no tendr n registros tipo "2" a continuaci n de su registro tipo "1". Para detectar aquellas viviendas sin personas se aconseja trabajar con registros de formato m ltiple, ya que con el formato  nico se crear  de todas maneras una entrada para una persona aunque las variables se consideraran con valores no aplica.

Es importante notar que el registro de vivienda tiene que venir ANTES de los registros de personas correspondientes. Es decir el archivo debe tener los registros ordenados en la forma vivienda-hogar- persona y no al revés. Se sugiere por ende marcar con "1" a las viviendas con un "2" a los hogares y con un "3" a las personas.

Solo el programa **CSPro trabaja con estructura de registros múltiples.**

El archivo de formato de registros múltiple es algunas veces llamado archivo "jerárquico", porque existe una cierta "jerarquía" entre los registros tipo "1" y "2", esto es, los registros tipo "2" "dependen" del registro previo tipo "1". Sin embargo, este no es realmente un archivo jerárquico, porque la parte de identificación (generalmente la información geográfica) es duplicada para todos los registros. Un verdadero archivo jerárquico tendría un tipo de registro por cada nivel de información (entidad), como en la tabla 5.

Tipo-reg Entidad

1	Provincia
2	Comuna
3	Distrito
4	Vivienda 1
5	Persona 1
5	Persona 2
5	Persona 3
4	Vivienda 2
5	Persona 1

Tabla 5. Representación de un archivo múltiple jerárquico

Cada entidad tendría un tipo de registro específico para almacenar sus datos. La información "geográfica" no sería repetida para cada registro en el archivo, pero sería almacenada sólo en el tipo de registro específico. Por ejemplo, la información de provincia (como su código, nombre u otra información de ese nivel) sería almacenada sólo en registros tipo "1".

Los archivos de registros múltiples no son manejados por muchos programas, (mencionamos aquí CSPro). Sin embargo, este tipo de archivo es muy útil para ahorrar espacio.

II. Creando una base REDATAM

II.1 ETAPAS PRELIMINARES

Una base de datos en formato REDATAM puede tener una estructura muy sencilla, con sólo una rama de información, o tan compleja como pueda imaginarse, con múltiples ramas en el árbol de la jerarquía. Puede ser muy pequeña en tamaño (sólo unos cuantos registros), o muy grande, incluyendo datos con cientos de registros. Obviando cuál es su caso, antes de comenzar el proceso de generación, hay que tomar en consideración dos cosas (1) garantizar que los datos de entrada están en un formato aceptado pro REDATAM y (2) verificar que los datos de entrada estén ordenados según la estructura deseada.

Un paso preliminar para generar una base de datos en REDATAM es analizar el contenido de los archivos de datos originales, la comprobación de su consistencia (por ejemplo, que no existan casos duplicados con el mismo código de identificación, que no existan hogares sin un jefe de hogar, etc.). Este paso es fundamental para el éxito del proyecto, ya que, si los datos originales no muestran la frecuencia requerida y / o son inconsistentes, la calidad de la base de datos se verá comprometida

REDATAM puede leer archivos en los formatos de CSPPro (dat y dcf), CSV (comma Separated Values), Dbase (dbf) o SPSS (sav). Si su archivo de entrada está almacenado en un tipo especial de formato, tales como aquellos usados en paquetes comerciales, por ejemplo, EXCEL, SAS, Oracle, SQL-Server, etc., se tienen que convertir todos los archivos de la base de datos (datos y diccionario) a uno de los formatos aceptados.

CONSIDERACIONES BÁSICAS PARA LA GENERACIÓN DE LA BASE DE DATOS

La generación de una base de datos en el módulo Red7 Create implica tres pasos básicos: (i) la definición de la estructura jerárquica; (ii) la definición de variables para el nuevo diccionario; y (iii) la carga de la base de datos misma. A continuación se explica en detalle cómo funciona cada paso. Estas definiciones se guardan en un esquema correspondiente a un archivo con la extensión .wipX. Este esquema es el paralelo del diccionario que se usara en la base final, es decir, se utiliza como referencia para la generación de la base de datos en formato REDATAM.

PASOS PARA LA GENERACION

La creación de una base de datos en formato REDATAM involucra los siguientes aspectos de manera secuencial:

1. Seleccionar el archivo original, verificando que su formato sea uno de los aceptados por *Redatam7*.
2. Análisis visual del archivo original para determinar las entidades y variables.
3. Analizar todos los campos del archivo original para determinar la estructura.
4. Abrir el programa Red7 Create.
5. Crear la jerarquía.
 - a. Nombrar la entidad raíz teniendo en cuenta el objetivo de los datos, con un tamaño recomendado de 8 caracteres.
 - b. Utilizar nombres de entidades que sean fácil de identificar por parte del usuario.
6. Guardar el Esquema de Creación (.wipX).
7. Abrir la definición del archivo de datos que corresponda.
8. Traspasar cada uno de los campos a su entidad respectiva.
9. Controlar que todos los campos se hayan pasado a la jerarquía (mostrar campos referidos).
10. Definir las entidades seleccionables (seleccionar la variables STRING que contiene los códigos de la entidad), es recomendable establecer como seleccionables todas las entidades -si se desea generar una base de uso interno y/o verificar la consistencia de los códigos de identificación-.
11. Utilizar la opción de “Verificar Valores” y/o “Verificar Máximo” para cada una de las entidades, guardar los archivos resultantes.
12. Establecer los valores mínimo y máximo (rangos) para las variables de tipo INTEGER y REAL, con ayuda de la información producida en el Paso 11.
 - a. Utilizar como valor 0 para NA en el caso de que no se utilice en el rango de valores válidos.
 - b. Para el valor MV (omitido) utilizar el valor máximo más uno.
 - c. En caso de existir valores para la categoría 0, verificar existencia de valores NA.
13. Identificar las variables numéricas que están definidas como STRING y convertirlas a INTEGER, asignando el rango con ayuda de la información del Paso 11.
14. Verificar el tamaño de las variables de códigos de las entidades (posición relativa y tamaño).

15. Documentar (rótulos para entidades, rótulos y categorías para variables).
 - a. En la entidad raíz, el rótulo es el nombre de la estadística que estamos creando.
 - b. En la entidad raíz incluir información sobre fecha, autor y lugar de Creación de la base de datos.
 - c. Grupo y Alias de las Variables.
16. Guardar el Esquema de Creación para actualizar los cambios.
17. Ejecutar la primera generación de la base de datos.

Paso 1 – Definir parámetros de la base de datos:

 - a. Nombre de la base de datos (nombre físico de los archivos .ptr y .rbf).
 - b. Nombre del diccionario *Redatam7* a utilizar con los otros módulos (Process, xPlan, Web, etc).
 - c. Localización de la base de datos REDATAM en un directorio diferente al de los archivos originales o de creación (.wipX, .lsdX).

Paso 2 – Definir el archivo de entrada.

Paso 3 – Seleccionar entidades, y tildar la verificación de la secuencia de datos en cada una de ellas.

Paso 4 – Seleccionar variables.

Seleccionar todas las variables para cada entidad (seccionar todo de todo)

Terminar – Generar la base de datos.

 - a. Guardar el esquema de generación (archivo .lsdX) en el directorio donde se encuentra el esquema de creación (archivo .wipX).
 - b. Ejecutar.
18. Examinar el informe de generación.
 - a. Comparar con el paso 2 (número total de entidades y variables).
 - b. Valores válidos, NA y MV (omitidos).
 - c. Verificar que el total de registros de la última entidad corresponda al total de registros del archivo original.
19. En el caso de existir inconsistencias o errores en el informe, hacer los cambios en el esquema de creación (.wipX) y/o en el archivo original.
20. Antes de volver a ejecutar, borrar el contenido del directorio donde se

- genera la base de datos REDATAM.
21. Ejecutar la generación abriendo el esquema de generación previamente creado (Paso 17).
 22. Verificar el informe de generación (Paso 18).
 23. Ejecutar los pasos 19 al 22 hasta llegar a obtener la base de datos REDATAM correcta, optimizada y documentada.
 24. Guardar el último informe de generación en el directorio donde se encuentre el esquema de creación (archivo .wipX).
 25. Ejecutar una frecuencia de todas las variables de la base con Red7 Process.
 26. Verificar las frecuencias de las variables y en el caso de inconsistencias volver al paso 19.
 27. Respalidar la base de datos REDATAM.
 28. Respalidar los archivos utilizados en la creación de la base de datos:
 - a. Archivo de datos.
 - b. Esquema de creación (.wipX).
 - c. Esquema de generación (.lsdX).
 - d. Informe de generación (.xls).

NOTA:

En Redatam7 Create existe libertad en asignar los nombres a entidades y variables, además se puede asignar alias o segundos nombres a cada entidad o variable para mantener el nombre original y poder utilizar un segundo nombre. Sin embargo, hay ciertas reglas a cumplir:

1. Los nombres pueden tener cualquier largo, se sugiere no sobrepasar 10 caracteres.
2. Pueden contener mayúsculas y minúsculas
3. Se sugiere no usar caracteres especiales (como @, #, \$, _), aunque éstos están permitidos.
4. No pueden usarse caracteres en blanco (espacios), en ninguna circunstancia.


II.2 REVISION DE LA CORRELACION GEOGRAFICA DE LA BASE DE DATOS

La revisión de la condición geográfica de una base de datos es la verificación de la secuencia de las entidades, es decir, la conexión o pertenencia en la jerarquía.

Como se dijo, el código geográfico está compuesto por una secuencia específica para cada uno de los elementos de una entidad. Cada parte del código informa la pertenencia a una determinada entidad, razón por la cual también se conoce como código compuesto. Esta codificación se realiza en el proceso de generación de la base, por lo cual se recomienda crear una base preliminar sólo con la estructura jerárquica y las variables principales de cada entidad para realizar una verificación.

Siguiendo los pasos identificados en la sección anterior, se crea la jerarquía, de acuerdo con las variables geográficas presentes. En cada entidad se agregará su variable seleccionable y en las entidades de vivienda, hogar y/o personas, como no puede ser identificable, solo se agregan las variables principales (Tipo de vivienda, Número de hogares, Sexo, Edad, Parentesco, etc.).

Una vez creada la jerarquía, se genera la base de datos preliminar, recordando tildar la verificación de la secuencia de datos en cada entidad seleccionable; ya que esto permitirá la revisión de la base. Ahora, para la revisión:

1. Se verifica en el informe de generación de la base, el número de entidades creadas y la secuencia entre ellas. Importante verificar los archivos punteros, ya que mostrarán también la secuencia numérica de la jerarquía (la pertenencia).
2. Verificar el número de registros por entidad y variable, que sea coincidente con la base de datos original, es decir, 5 departamentos, 31 provincias, 95 municipios, etc.
3. Abrir la nueva base de datos en el módulo Red7 Process para verificar la jerarquía creada, y en cada entidad, que contenga las variables correspondientes al código geográfico se acompañan del símbolo .
4. Una revisión rápida para ver la secuencia es realizar una lista por área de una de las variables principales agregadas a esta base preliminar. Para eso, en la pestaña "Procesador estadístico" seleccionar el proceso "Lista por Área". Se despliega una nueva ventana para seleccionar el nivel de salida, el cual corresponde al nivel geográfico. Para este caso, se selecciona el nivel más bajo de la jerarquía, y luego se selecciona una variable para la tabla. Se ejecuta y se verifica la salida.

El resultado debe ser una tabla, donde la primera columna sea el REDCODE, es la construcción de la secuencia que corresponde al código geográfico, y luego las categorías de la variable seleccionada. Este RECODE es un agregado de los códigos de cada variable geográfica de cada entidad.

II.3 EJEMPLO DE CARGA CON DATOS CSV

Para la revisión de los datos de entrada, se ubica el archivo y se hace una revisión rápida de datos. Es un archivo que incorpora un signo o delimitador entre valores, siendo coma (,), punto y coma (;) o carácter invisible (tab) los más utilizados. En este tipo de datos se deben definir los rótulos, rangos de cada variable directamente en la creación de la base.

	A	B	C	D	E	F
1	DEPTO;PROVIN;MUNIC;DISTRITO;MANZANA;HOGAR;PERSONA					
2	1;07;010;0701;07;02;01;1;3;1;2;1;4;2;3;3;6;1;2;2;1;2;72;3;8;1					
3	1;07;010;0701;07;02;02;1;3;1;2;1;4;2;3;3;6;1;2;2;5;2;11;1;0;0					
4	1;07;010;0701;07;03;01;1;3;6;2;1;4;2;3;3;6;1;2;1;1;1;48;3;1;1					
5	1;07;010;0701;07;03;02;1;3;6;2;1;4;2;3;3;6;1;2;1;2;2;48;3;1;1					
6	1;07;010;0701;07;03;03;1;3;6;2;1;4;2;3;3;6;1;2;1;3;1;26;3;3;1					
7	1;07;010;0701;07;03;04;1;3;6;2;1;4;2;3;3;6;1;2;1;3;2;19;1;3;6					
8	1;07;010;0701;07;03;05;1;3;6;2;1;4;2;3;3;6;1;2;1;3;2;17;1;8;6					
9	1;07;010;0701;07;03;06;1;3;6;2;1;4;2;3;3;6;1;2;1;3;1;9;1;0;0					
10	1;07;010;0701;08;05;01;1;6;1;2;1;4;2;3;3;6;1;2;2;1;1;78;3;6;7					
11	1;20;010;1408;04;03;01;1;1;1;2;1;4;2;3;3;6;1;1;2;1;1;47;3;1;1					

Figura 4: Datos de entrada CSV

CREACIÓN DE LA ESTRUCTURA DE LA BASE DE DATOS

Al abrir el módulo Red7 Create y comenzar la creación del diccionario, se selecciona la opción "Nuevo" desde el menú principal, en la pestaña "Archivo". Una vez abierta la ventana en la parte izquierda del módulo, en donde se creará la jerarquía de la base; se debe realizar la conexión con los datos de origen.

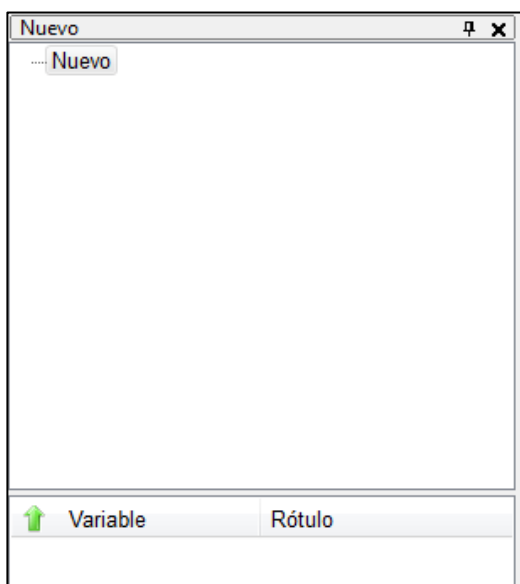


Figura 5: Ventana de creación de jerarquía base de datos, Red7 Create.

Para ello, en la pestaña “Datos de Origen” se selecciona bajo “Conectar” los datos de origen según su formato, y se habilita en la parte derecha del módulo una ventana con la información de los datos conectados.



Figura 6: Menú dinámico módulo Red7 Create

En la ventana derecha del módulo, se debe localizar el archivo de origen y definir el separador de campos que presenta dicho archivo. Una vez realizado, se selecciona “Cargar desde Datos” y se despliegan todas las variables disponibles.

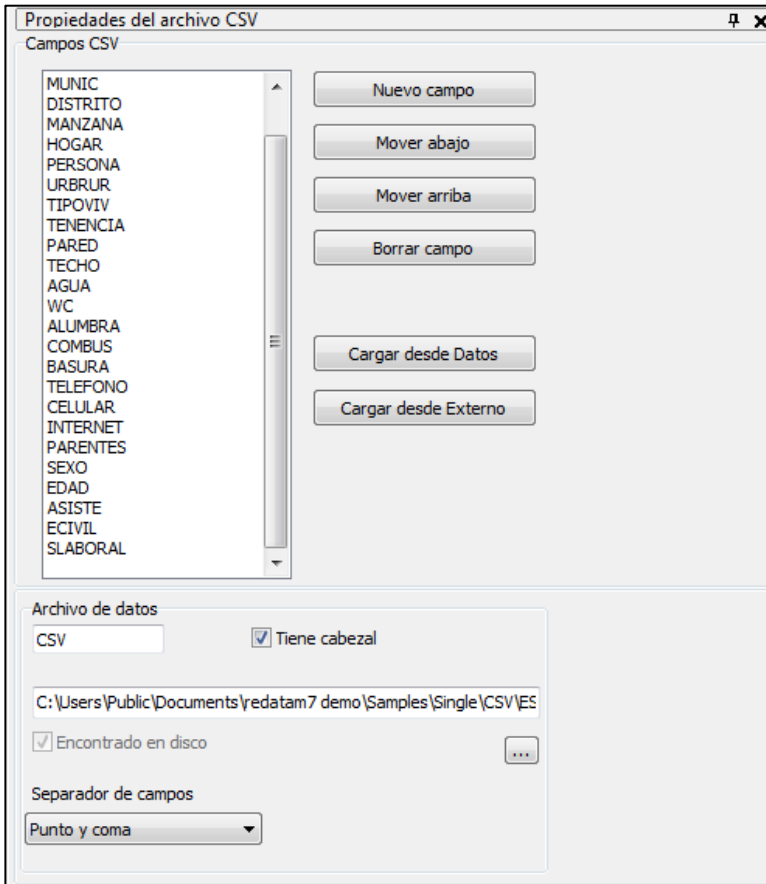


Figura 7: Ventana de propiedades de los datos de origen, formato CSV.

El siguiente paso es crear la jerarquía de la base, según las variables. Para ello se renombra la raíz que se habilitó al seleccionar “Nuevo” al abrir Create. Luego, se van

agregando las entidades sobre la última creada, presionando clic derecho sobre ella y la opción “Agregar entidad inferior” o “Insertar entidad hermana”, según sea el caso.

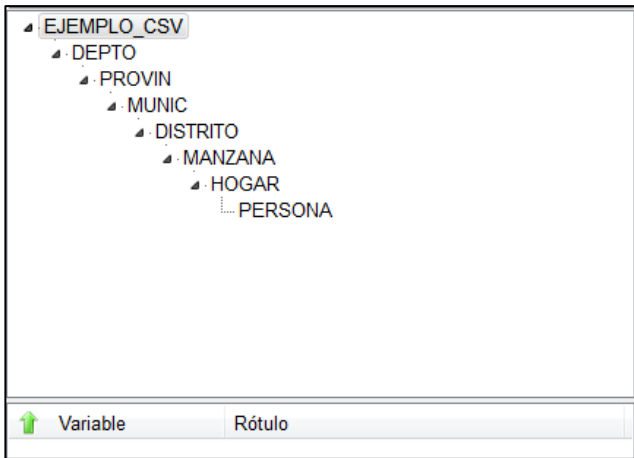
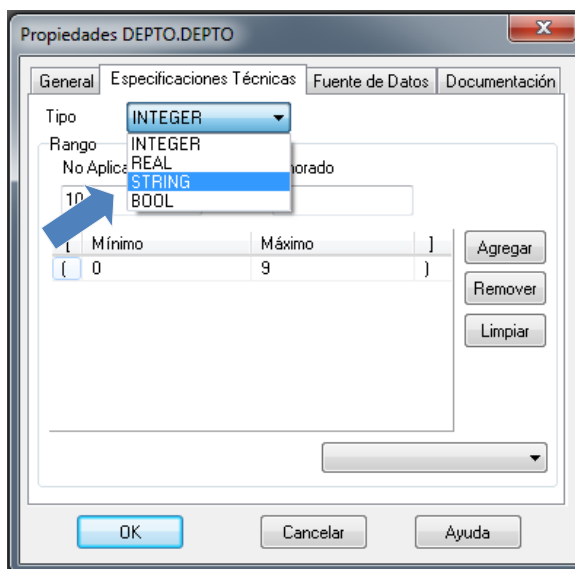


Figura 8: Jerarquía de la base de datos.

Seguidamente hay que trasladar las variables a la entidad que corresponda y para ello se siguen las siguientes pautas:

1. Primero se selecciona la entidad a la cual se le va a agregar uno o varios campos, y una vez en ella, se arrastran desde la ventana de la derecha hacia el sector inferior de la ventana izquierda las variables.
2. Para cada entidad se debe identificar cuál de las variables contiene el código identificador (ordenador) de la entidad. Esta variable debería ser de tipo carácter (String). Si esta variable no fuera de tipo carácter se podría modificar directamente en el esquema REDATAM. Para ello se hace presiona clic derecho sobre la variable y seleccionar propiedades o doble clic y se abre la ventana automáticamente; luego se modifica a STRING y al confirmar con Ok la variable cambia.



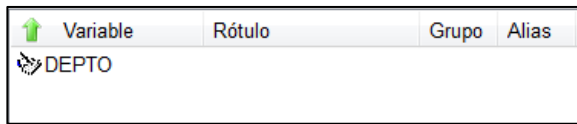


Figura 9: Ventana de propiedades de la variable

3. Una vez arrastrada la variable a cada entidad e identificada la variable con el código ordenador, se procede a definirla como código de la entidad, esto se hace: (1) abriendo la ventana de propiedades de la entidad DEPTO desde el menú dinámico (haga clic en el botón derecho del mouse sobre el nombre de la entidad DEPTO), y (2) en la ventana que se abre elija la variable DEPTO desde el casillero "Variable de códigos". Luego marque el casillero "Seleccionable" para distinguir a la entidad como entidad seleccionable. Repita este proceso para cada entidad de la estructura jerárquica y guarde el esquema.

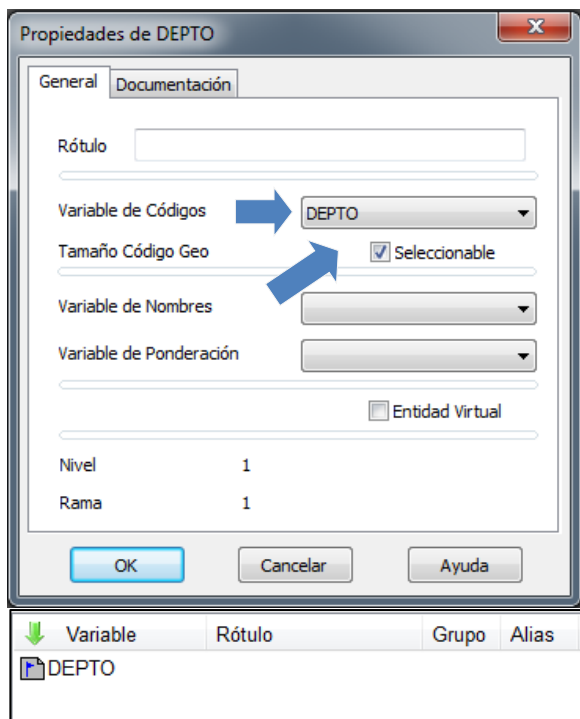


Figura 10: Ventana de propiedades de la entidad

4. Luego se procesa a arrastrar el resto de las variables a cada entidad, desde la ventana de origen hacia la ventana del esquema asignando.
5. Para cada variable hay que determinar las propiedades de la misma, ya que no existe diccionario en este formato. Para ello debemos contar con la boleta o cuestionario de recogida de la información, para verificar los rótulos y rangos. Los pasos para cada variable son los siguientes:
 - a. Sobre la variable se presiona botón derecho y se selecciona propiedades (o doble clic sobre la variable). Una vez en ella, en la pestaña "general" se edita el rótulo y se agregan las categorías y sus códigos. Luego en la

pestaña "Especificaciones técnicas" se edita el rango de la variable. Y en documentación se puede crear el Alias y Grupo a cada variable. El campo de "grupo" de la variable es usado solamente para despliegue, con el objeto de agrupar las variables pertenecientes a un mismo tema (un clic en el cabezal de la columna de grupos clasifica la lista de variables según su agrupación). El campo de "alias" de la variable puede ser usado en la programación de Redatam7 en lugar del nombre de la variable.

b. Al finalizar, se presiona Ok y la variable queda configurada.

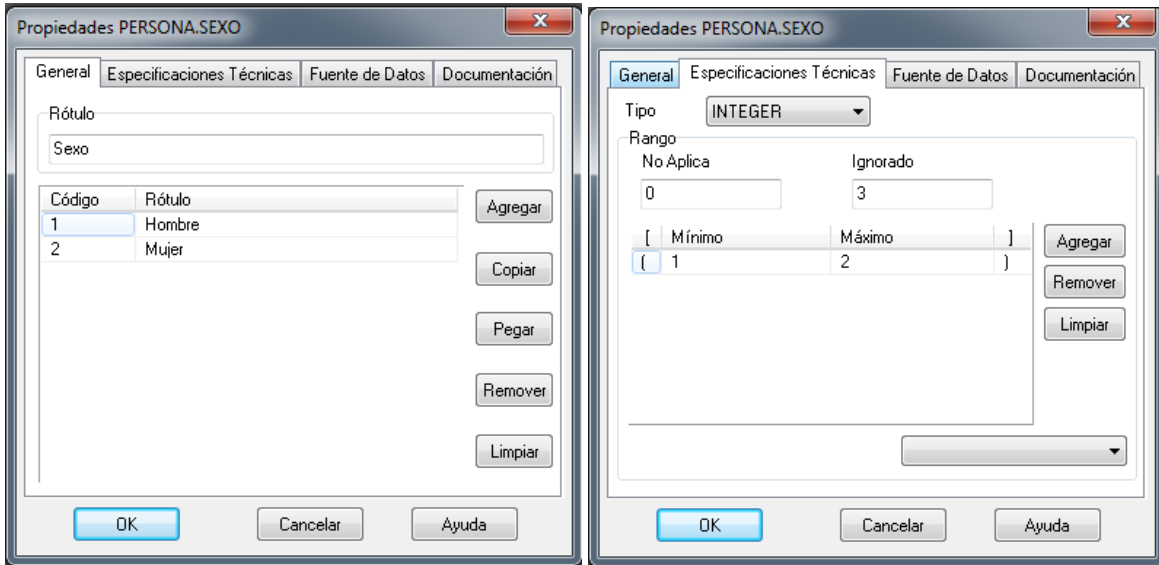


Figura 11: Ventana de propiedades de la variable.

6. El procedimiento se repite para cada variable de cada entidad. Y una vez configuradas, se hace una revisión final al esquema antes de la generación de la base para garantizar que todas las entidades tengan su código identificable y que las variables presenten sus rótulos, categorías y rangos.

II.4 EJEMPLO DE CARGA CON DATOS CSPRO

Antes de la creación de la base de datos, se debe realizar una revisión del archivo de entrada y sus datos. Se toma como ejemplo una base de entrada múltiple y se verifica que presenta variables comunes para dos secciones: Vivienda y Personas en la vivienda.

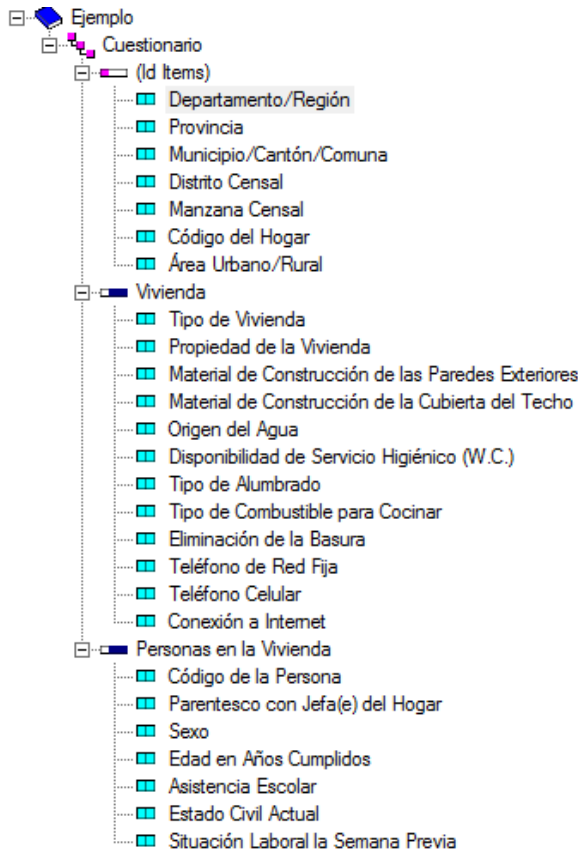


Figura 12: Formato de datos CsPro

A diferencia del ejemplo anterior, las variables presentan sus rótulos, categorías y valores.

N	Value Set Label	Value Set Name	Value Label	From	To	Special	Image
<input type="checkbox"/>	Tipo de Vivienda	TIPOVIV_VS1					
<input type="checkbox"/>			Casa	1			...
<input type="checkbox"/>			Departamento en edificio	2			...
<input type="checkbox"/>			Piezas en casa antigua o conventillo	3			...
<input type="checkbox"/>			Mejora, mediagua	4			...
<input type="checkbox"/>			Rancho, choza	5			...
<input type="checkbox"/>			Móvil (carpa, vagón, container, bote, lancha, similar)	6			...
<input type="checkbox"/>			Otro	7			...

Figura 13: Ventana de propiedades de la variable en formato CsPro

CREACIÓN DE LA ESTRUCTURA DE LA BASE DE DATOS

Entendiendo la configuración del diccionario, se procede a crear la jerarquía deseada. Para la creación del diccionario se selecciona la opción "Nuevo" desde el menú principal, para abrir un esquema o ventana nueva, en donde se definirá la estructura jerárquica. Luego bajo el menú datos de origen se selecciona bajo "conectar" los datos de origen y así importar el diccionario de los datos originales en formato CSPro.



Figura 14: Menú dinámico módulo Red7 Create

Una vez seleccionado, se despliega en el extremo derecho de la pantalla la con la información general del archivo de entrada. En la pestaña General (1), su tipo, nombre, ruta, y la conexión el diccionario; y en la pestaña Campo (2), todos los campos asociados a los datos de entrada.

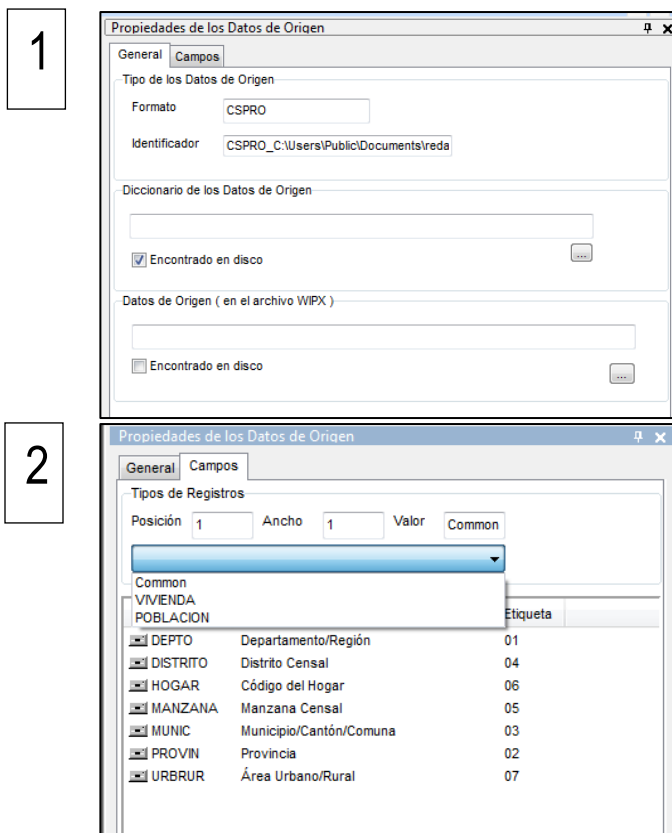


Figura 15: Ventana de propiedades de los datos de origen, formato CsPro.

Una vez establecida la conexión, hay que definir la jerarquía en la base Redatam y trasladar las variables a la entidad que corresponda para crear el diccionario. Para ello seguimos los siguientes pasos:

1. Primero se selecciona la entidad a la cual se le va a agregar uno o varios campos; así como la lista de registros a la cual pertenece las variables. Definido esto, se arrastran desde la ventana de la derecha hacia el sector inferior de la ventana izquierda.

2. Para cada entidad se debe identificar cuál de las variables contiene el código identificador (ordenador) de la entidad. Esta variable debería ser de tipo carácter (String). Si esta variable no fuera de tipo carácter se podría modificar posteriormente en el esquema REDATAM. Para ello se hace presiona clic derecho sobre la variable, propiedades, luego se modifica a STRING (1) y Ok y la variable cambia (2).

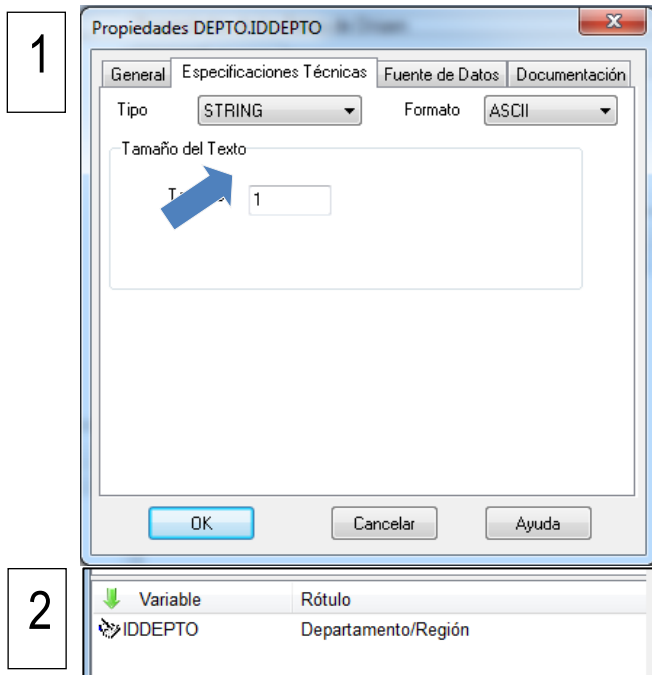


Figura 16: Acceso a la ventana de propiedades de la variable.

3. Una vez arrastrada la variable a cada entidad e identificada la variable con el código ordenador, se procede a definirla como código de la entidad, esto se hace: abriendo la ventana de propiedades de la entidad DEPTO desde el menú dinámico (haga clic en el botón derecho del mouse sobre el nombre de la entidad DEPTO), y (1) en la ventana que se abre elija la variable DEPTO desde el casillero "Variable de códigos". Luego marque el casillero "Seleccionable" para distinguir a la entidad como entidad seleccionable. Repita este proceso para cada entidad de la estructura jerárquica y guarde el esquema.

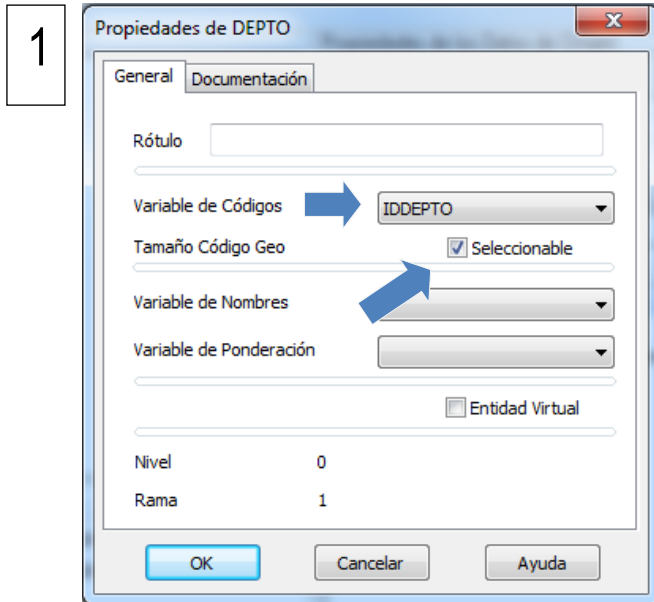


Figura 17: Ventana de propiedades de la entidad

4. Luego se procesa a arrastrar el resto de las variables desde la venta de origen hacia la ventana del esquema asignando la o las variables a cada grupo o entidad (las variables de vivienda o de personas según corresponda).
5. Luego se revisa cada variable a través de la ventana de propiedades (clic derecho sobre la variable) y se verifica el rotulo y categorías, tipo de variable y rangos, creación de alias y/o grupo para cada variable.

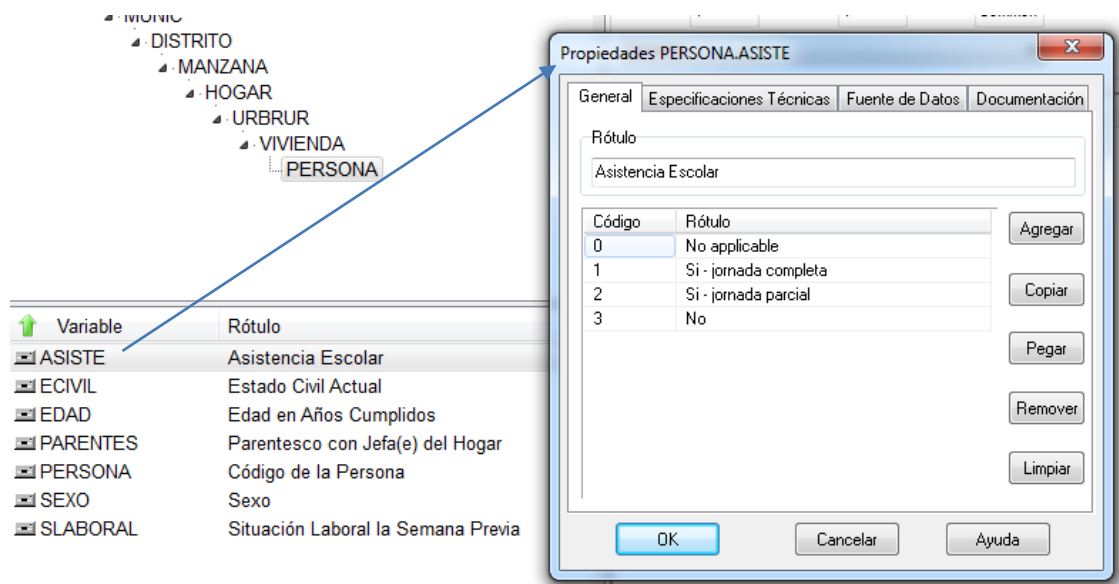


Figura 18: Verificación de las propiedades de la variable.

GENERACIÓN DE LA BASE DE DATOS

A continuación, se procede a generar la base de datos REDATAM. Para esto, seleccione la opción “Ejecutar Creación” desde el menú “Base de Datos

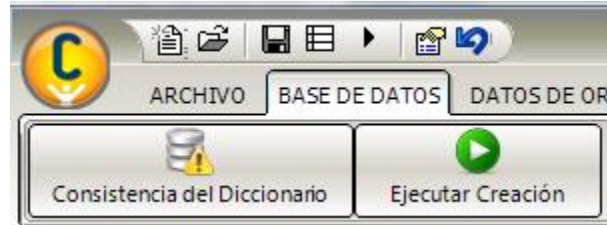


Figura 19: Menú dinámico de módulo Red7 Create, para la generación de la base.

Se despliega una nueva ventana para iniciar el proceso de generación de la base de datos. En ella es necesario definir todos los parámetros que se incluyen en cada una de las páginas de esta ventana.

1. El primer paso consiste en definir un nombre de la base de datos a generar, el directorio en el cual se van a localizar los archivos de variables (.rbf), entidades (.ptr) y diccionario (.dicX).

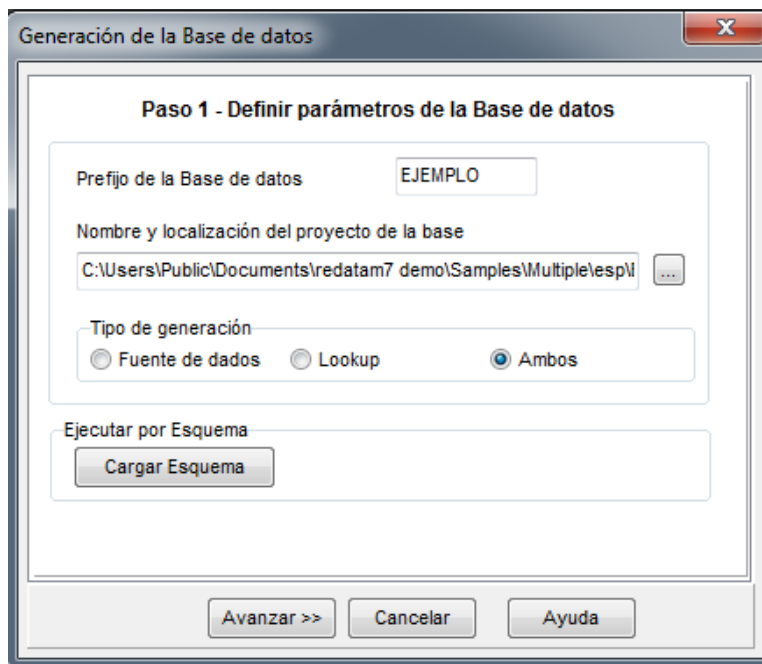


Figura 20: Ventana paso 1: localización archivo salida.

2. El segundo paso consiste en definir el formato y archivo de entrada de los datos.

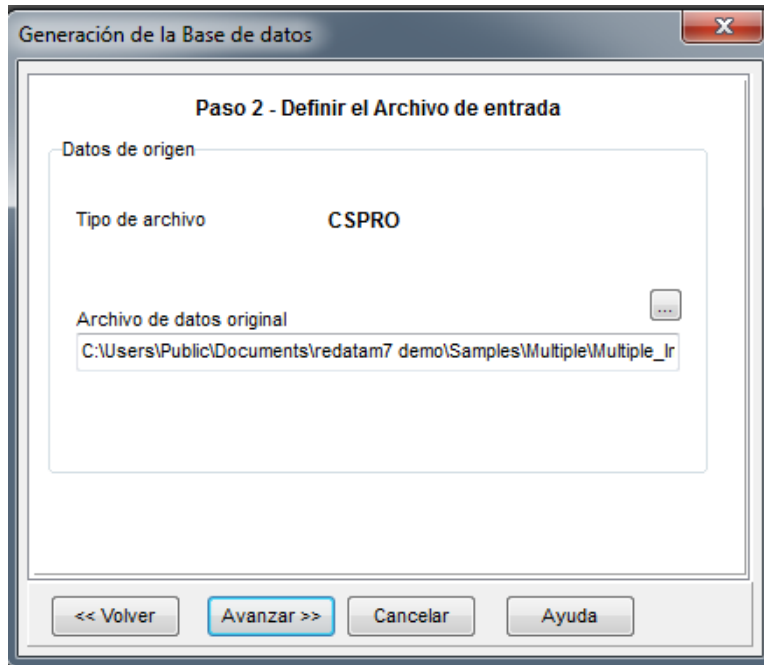


Figura 21: Ventana paso 2:localización archivo de entrada.

3. El tercer paso consiste en definir para cada entidad si se crea o no un archivo de punteros y generar el **código compuesto** asociado a la entidad; también permite verificar la secuencia de los registros según el código geográfico identificador, para lo cual es necesario que esté activado el recuadro "Verificar Secuencia de Datos".

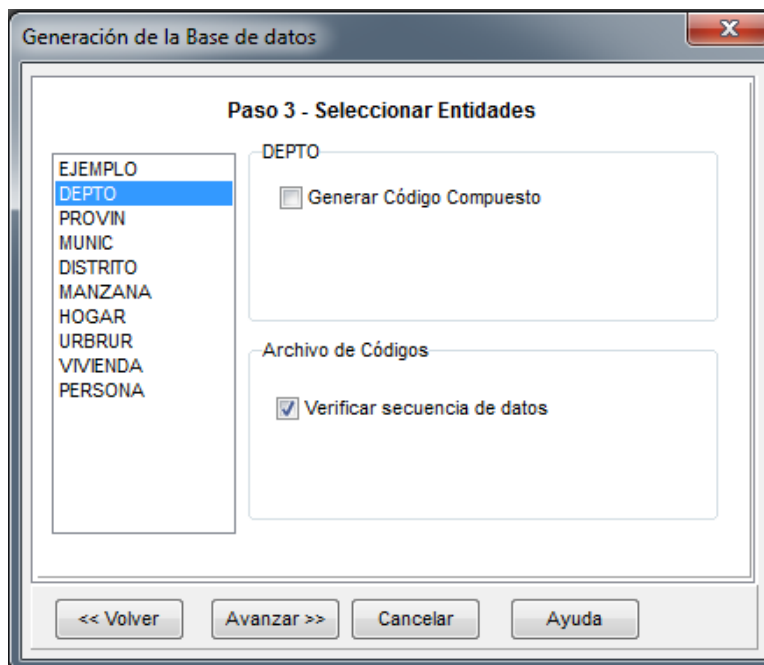


Figura 22: Ventana paso 3: selección de entidades.

NOTA: Para el caso de una base múltiple (Ejemplo de CsPro), se debe seleccionar el “Quiebre según tipo de registro” la sección de la cual serán tomadas las variables para las entidades que corresponda y que no sean de uso común.

4. En el cuarto paso, se debe definir las variables a incluir en la generación de la base de datos en formato REDATAM para cada entidad.

Utilizando el mouse seleccione la entidad que contiene las variables y marcando en cada casillero para elegir las variables o en el botón “Seleccionar Todo” en el caso de todas las variables de la entidad seleccionada. Adicionalmente tiene la opción de “Seleccionar Todo de Todo” para marcar todas las variables existentes en el esquema de creación.

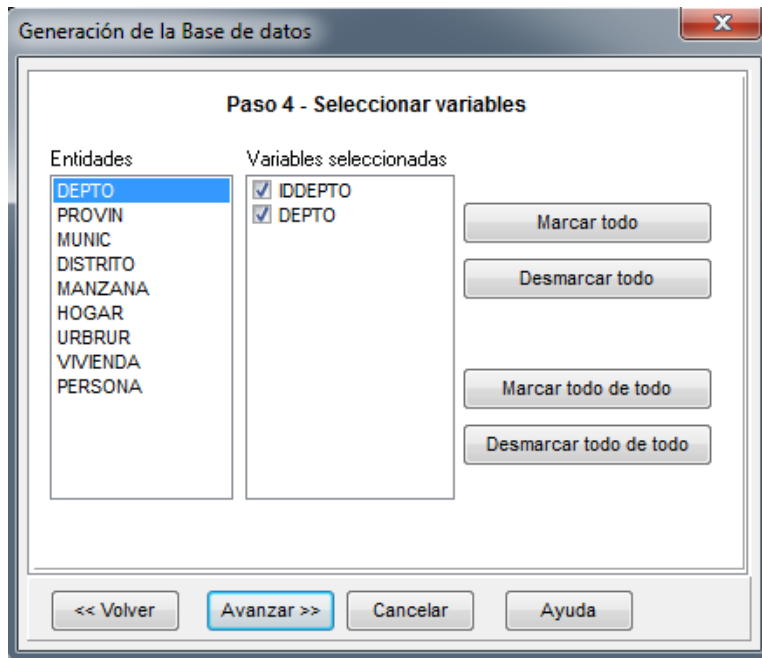


Figura 23: Ventana paso 4: selección de variables

5. En el quinto paso y final, se puede almacenar los parámetros definidos en los pasos anteriores en un archivo especial (*.IsdX) con la opción "Guardar Esquema", para no tener que repetirlos en una generación posterior de la misma base de datos.

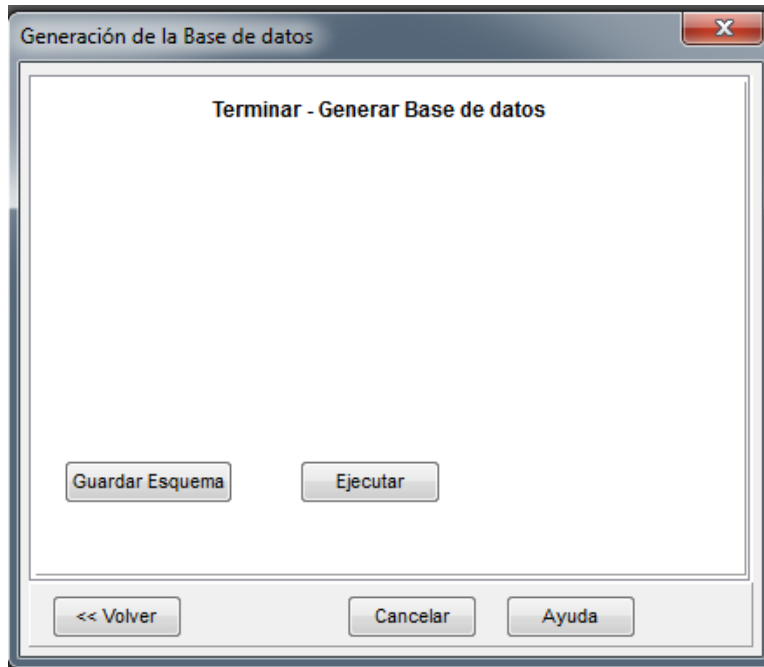


Figura 24: Ventana paso 5: generación de base de datos

6. Finalmente, concluir el proceso de generación utilizando el botón "Ejecutar". Una vez terminado el proceso de generación de la base de datos, la ventana de ejecución se cierra y se despliega el Informe de Generación, el cual debe ser revisado minuciosamente para detectar posibles inconsistencias como: valores fuera de rango, valores no aplica, total de registros para cada entidad, etc.

Debe mencionarse que todas las variables están comprimidas en formato binario. Es decir, la información correspondiente a las variables se ha almacenado en archivos codificados (.rbf), uno por variable. La información correspondiente a la pertenencia de una entidad a otra entidad superior se ha almacenado en los archivos de punteros (.ptr), uno por entidad.

La definición de las variables y las entidades se ha almacenado en un archivo de definición (.wipX), así como también se ha generado un archivo con el esquema de generación (.lsdX) y no olvidemos el Informe de Generación (.xls).

II.5 REVISION DEL INFORME DE GENERACION

El informe de generación es un archivo en formato Excel, que contiene la información de las entidades y sus variables. Si bien es un solo archivo, se puede analizar en dos partes, siendo la segunda la más importante:

1. Información general: se refiere a los archivos usados y creados, y su localización.

Generación de bases de datos Red7 - Informe de generación				
Reporte de generación guardado en			C:\Users\Public\Documents\redatam7 demo\Samples\Multiple\esp\ReporteGeneracion.xls	
Fecha	Inicio	Término	Duración	Estado
9/11/2018	11:33:47	11:33:48	00:00:00	Ha sido generado con éxito
Diccionario de los datos de origen			C:\Users\Public\Documents\redatam7 demo\Samples\Multiple\esp\Multiple_Layout_Esp.dcf	
Datos de origen			C:\Users\Public\Documents\redatam7 demo\Samples\Multiple\Multiple_Input.dat	
Red7 Estructura de generación			C:\Users\Public\Documents\redatam7 demo\Samples\Multiple\EJEMPLO.wipX	
Red7 Esquema de generación			C:\Users\Public\Documents\redatam7 demo\Samples\Multiple\esp\EJEMPLO.lsdX	
Red7 Directorio de la base de datos			C:\Users\Public\Documents\redatam7 demo\Samples\Multiple\esp	
Red7 Proyecto			EJEMPLO.prjX	
Red7 Diccionario			.BaseR\EJEMPLO.dicX	

Figura 25: Informe de generación.

2. Información específica: referida a la información de cada entidad y variable. Indica el número de entidades y variables creadas, el número de registros válidos para cada una, así como el número de registros de “no aplica” o “inválido”; y la localización del puntero (.ptr archivo de conexión de la jerarquía) y del archivo de datos (.rbf archivo con información de la variable).

Lo más importante a verificar en el informe es el número de registro por cada entidad y variable. Los registros son inversamente proporcionales a la jerarquía de la base, es decir, mientras más alejada de la raíz este la entidad o variable, mayor será el número de registros.

Número	Entidad/Variable	Registros/Valores	No se aplica	Inválido	Puntero/Archivo de Datos/Filtro
1	EJEMPLO	1			.\BaseRJEJEMPLO_EJEMPLO.ptr
1_1	EJEMPLO	1	0	0	.\BaseRJEJEMPLO_EJEMPLO.rbf
1_2	REDCODEC	1	0	0	.\BaseRJEJEMPLO_EJEMPLO_REDCODEC.rbf
2	DEPTO	5			.\BaseRJEJEMPLO01.ptr
2_1	IDDEPTO	5	0	0	.\BaseRJEJEMPLO01.rbf
2_2	DEPTO	5	0	0	.\BaseRJEJEMPLO02.rbf
3	PROVIN	31			.\BaseRJEJEMPLO02.ptr
3_1	IDPROVIN	31	0	0	.\BaseRJEJEMPLO03.rbf
3_2	PROVIN	31	0	0	.\BaseRJEJEMPLO04.rbf
4	MUNIC	35			.\BaseRJEJEMPLO03.ptr
4_1	IDMUNIC	35	0	0	.\BaseRJEJEMPLO05.rbf
4_2	MUNIC	35	0	0	.\BaseRJEJEMPLO06.rbf
5	DISTRITO	133			.\BaseRJEJEMPLO04.ptr
5_1	IDDISTRITO	133	0	0	.\BaseRJEJEMPLO07.rbf
5_2	DISTRITO	133	0	0	.\BaseRJEJEMPLO08.rbf
6	MANZANA	577			.\BaseRJEJEMPLO05.ptr
6_1	IDMANZANA	577	0	0	.\BaseRJEJEMPLO09.rbf
6_2	MANZANA	577	0	0	.\BaseRJEJEMPLO10.rbf
7	HOGAR	665			.\BaseRJEJEMPLO06.ptr
7_1	IDHOGAR	665	0	0	.\BaseRJEJEMPLO11.rbf
7_2	HOGAR	665	0	0	.\BaseRJEJEMPLO12.rbf
7_3	URBRUR	665	0	0	.\BaseRJEJEMPLO13.rbf
7_4	TIPOVIV	665	0	0	.\BaseRJEJEMPLO14.rbf
7_5	TENENCIA	665	0	0	.\BaseRJEJEMPLO15.rbf
7_6	PARED	665	0	0	.\BaseRJEJEMPLO16.rbf
7_7	TECHO	665	0	0	.\BaseRJEJEMPLO17.rbf
7_8	AGUA	665	0	0	.\BaseRJEJEMPLO18.rbf
7_9	WC	665	0	0	.\BaseRJEJEMPLO19.rbf
7_10	ALUMBRA	665	0	0	.\BaseRJEJEMPLO20.rbf
7_11	COMBUS	665	0	0	.\BaseRJEJEMPLO21.rbf
7_12	BASURA	665	0	0	.\BaseRJEJEMPLO22.rbf
7_13	TELEFONO	655	10	0	.\BaseRJEJEMPLO23.rbf
7_14	CELULAR	647	18	0	.\BaseRJEJEMPLO24.rbf
7_15	INTERNET	625	40	0	.\BaseRJEJEMPLO25.rbf
8	PERSONA	2214			.\BaseRJEJEMPLO07.ptr
8_1	PARENTES	2214	0	0	.\BaseRJEJEMPLO26.rbf
8_2	SEXO	2214	0	0	.\BaseRJEJEMPLO27.rbf
8_3	EDAD	2214	0	0	.\BaseRJEJEMPLO28.rbf
8_4	ASISTE	2214	0	0	.\BaseRJEJEMPLO29.rbf
8_5	ECIVIL	1580	634	0	.\BaseRJEJEMPLO30.rbf
8_6	SLABORAL	1580	634	0	.\BaseRJEJEMPLO31.rbf
8_7	PERSONA	2214	0	0	.\BaseRJEJEMPLO32.rbf

Figura 26: Informe de generación.

También es importante destacar el total de registros para una variable es la suma de los valores válidos, los no aplica y los inválidos. Es así que el total de una variable no puede ser mayor al total de registros de esa entidad.

Por ejemplo: Para Hogar se tienen 665 registros, y la mayoría de sus variables tienen el mismo número de registro; sin embargo, la variable “teléfono” presenta

solo 655 registros válidos, pero al sumar los registros de “no aplica” (10 registros), resulta el total de registros de la entidad. Para el caso de personas es similar. La entidad tiene un total de registros válidos de 2214, sin embargo, la variable “ecivil” presenta solo 1580, pero al sumar los “no aplica” resulta el total válido.

III.Revisión de la base de datos con Redtama7 Process

CONSISTENCIA DEL CONTENIDO DE LA BASE DE DATOS

Comprendiendo la composición de la base de datos, se puede verificar su contenido. Esto se realiza en el módulo Redatam7 Process. Para ello, se debe tener en cuenta las siguientes observaciones:

- Es necesario contar con las boletas censales o cuestionarios, según sea el caso, para verificar rangos, rótulos, saltos de preguntas, etc.
- El universo de la base o total corresponde a la cuantía de todos los casos o registros que componen cada entidad y se puede verificar de dos maneras: posicionando el cursor en una entidad determinada (por ejemplo, Viviendas) y presionando la tecla f8, lo que despliega una ventana en la que se puede observar o leer el número de elementos pertenecientes a la entidad o recurriendo a variables (o preguntas) que obligatoriamente deben ser respondidas o llenadas en cualquier proceso (Censo o Encuesta). Cada entidad debe tener una variable de este tipo y debe ser identificada para la evaluación de la base. Por ejemplo, para Personas las principales son Sexo y Edad; en el caso de Viviendas puede ser Tipo de Vivienda.
- El total de una variable con saltos es la suma de los valores “válidos” más los valores de “no aplica” y los “inválidos”.
- Para la verificación de consistencia de cada variable se toma como proceso principal FREQUENCY, a fin de verificar totales en las mismas.

REVISAR LA BASE DE DATOS

Una vez abierto el diccionario, se muestra la estructura y las variables, es decir, los niveles o entidades en que se organiza la información, dibujándose en forma de árbol la jerarquía que comanda la organización de los datos y las variables asociadas a ellas.

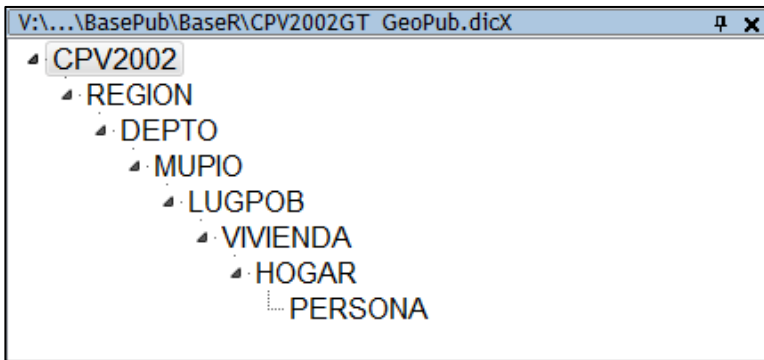





Figura 27: Diccionario del Censo Guatemala 2002.

- Cada variable original de la base de datos se identifica con el símbolo .
- Aquellas variables correspondientes al código geográfico se acompañan del símbolo .
- Aquellas variables que contienen los nombres de entidades se acompañan del símbolo .

Se puede seleccionar una variable (ubicando el puntero del mouse y haciendo "doble clic" sobre ella) y revisar algunas características como rótulo, categorías, rangos mínimo y máximo, disponibilidad de la variable y alguna otra información que le pudiese ser de utilidad.

PROCESOS DE FRECUENCIAS Y UN PLAN DE TABULADOS

La frecuencia es un proceso que tabula la distribución de datos de una variable. Esta distribución permite observar los registros válidos, no aplica y missing; es por ellos que es una de las principales maneras de evaluar la consistencia de la base. Existen tres formas de hacer la frecuencia.

Hacer una revisión de las variables y entidades, una vez fue verificada la consistencia de la base de datos, es el siguiente paso.

Un plan de tabulados se refiere a la programación de una serie de frecuencias para las variables de una base de datos. Es la escritura de una sintaxis genérica, utilizando los nombres de las variables, sus alias o grupos, para generar una serie automática de frecuencias. El plan se puede mantener como predefinido, ameritando solo la adaptación de los nombres de variables de acuerdo con la base de datos que se esté trabajando. Para el caso de los censos, trabajar con los Alias o Grupos, permite conservar el modelo de plan de tabulado.

En cualquiera de los casos, la sintaxis debe estar escrita tal cual el nombre del alias o grupo en la base, es decir, conservar mayúsculas y minúsculas, o de las variables requeridas.

IV. Módulo Redatam7 Admin para modificar una base de datos

La familia REDATAM presenta el módulo ADMIN para el manejo y administración de bases de datos. Permite la conversión entre los distintos diccionarios de Redatam, e incluye, además, facilidades para expandir, concatenar, descargar, comparar, modificar y otras acciones, para las bases de datos Redatam.

CORRECCIONES DE VARIABLES

Una base de datos o diccionario puede tener algunos errores o diferencias en los rótulos y categorías. Si bien los datos ya vienen creados con rangos específicos y cambiar una variable técnicamente no es posible, pues viene con especificaciones establecidas, lo cual ameritaría la creación de una variable derivada de ésta; es posible realizar ediciones a los rótulos y categorías de la variable original.

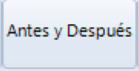
Para ello, en el módulo de ADMIN, en la misma sección de propiedad de la variable (botón derecho sobre la variable), se pueden hacer las ediciones pertinentes de cada variable; siempre recordando guardar el diccionario una vez hechas las modificaciones.

ANTES Y DESPUÉS

El “Antes y Después” es uno de los utilitarios del módulo Admin y su objetivo es comparar dos bases de datos que tengan la misma estructura y número de casos, a través de la combinación de ambas se facilita la comparación de los archivos en dos períodos, antes y después, de los programas de consistencia.

Para esta comparación, el programa crea un diccionario combinado a partir del archivo de entrada y el archivo de salida, asignando una letra inicial a cada variable, según el archivo. El hecho de que exista un par de variables (antes y después) para cada variable de entrada posibilita la comparación de ellas, y el estudio y verificación de los cambios efectuados por los procesos de consistencia y asignación.

Para su uso, se debe abrir el módulo ADMIN y en la pestaña *Utilitarios*, se presiona el

botón  y se despliega una ventana con las siguientes entradas:

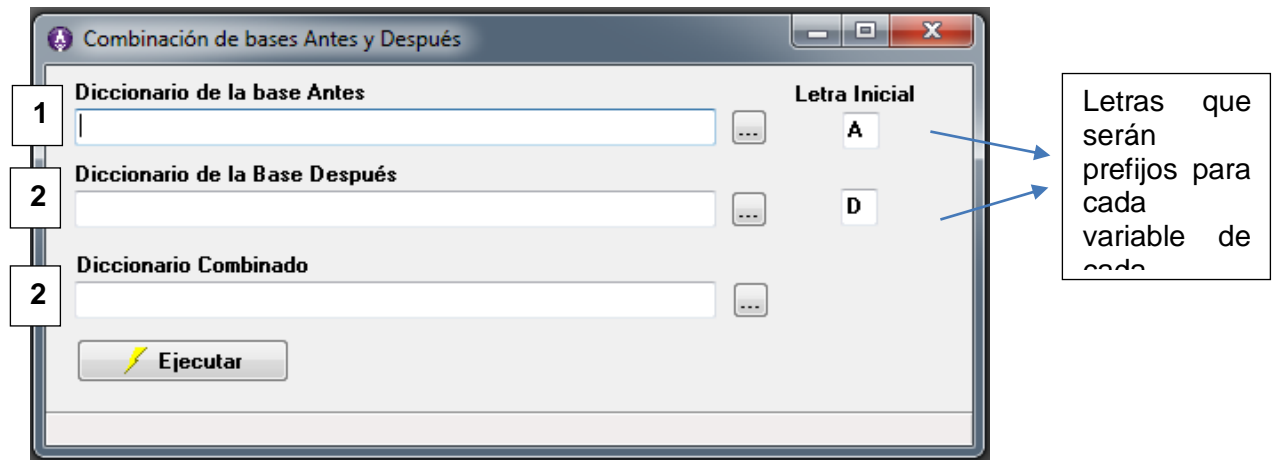


Figura 28: Ventana de “Antes y Después”, módulo ADMIN.

1. Ruta del diccionario del antes (original)
2. Ruta del diccionario del después (modificado)
3. Nombre y ruta del nuevo diccionario combinado

Antes de utilizar esta función, es importante destacar las siguientes consideraciones:

- Las bases de entrada deben tener la misma estructura y mismos nombres de entidades.
- Las bases de entrada deben tener el mismo número de casos en cada entidad.
- El número de variables y sus nombres NO tienen que ser los mismos en las bases de entrada.
- Las letras de prefijo a ser asignadas a las variables tienen que ser únicas (una sola letra) y no pueden ser iguales entre sí. Sin embargo, una de ellas puede ser en blanco (sin prefijo).
- Se recomienda que los nombres de los diccionarios sean distintos, sobre todo el diccionario de salida.

Una vez ejecutada la comparación, se genera el nuevo diccionario con todas las variables, diferenciadas por la letra -prefijo- asignada. Este nuevo archivo posibilita la comparación de cada par de variable, verificando los cambios efectuados por los procesos de consistencia previos.

Para entender mejor su utilidad, se toma de ejemplo la variable SEXO. El nuevo diccionario tendrá dos variables: ASEXO y DSEXO. Con dichas variables se pueden hacer los siguientes procesos:

- a) Cruce de ASEXO por DSEXO para observar todos los cambios entre las categorías (ignorado → hombre, ignorado → mujer, hombre → mujer y mujer → hombre).
- b) La creación de una variable temporal para cada par de variables, con lo cual se logra evaluar si hubo cambio en sus valores. Esa variable tendría dos categorías (0 – sin

cambios, y 1 – con cambios), y con una frecuencia simple de ellas, se evidencia rápidamente el número de cambios en cada una de las variables originales.

```
DEFINE PERSONA.EVASEXO
AS PERSONA.ASEXO <> PERSONA.DSEXO
TYPE INTEGER
DEFAULT 0
RANGE 0:1
VALUETAGS
0 "Sin cambios en la variable"
1 "Con cambios en la variable"
```

- c) La suma de todas las variables temporales entrega el número de variables asignadas por persona. La frecuencia de esta variable informa cuántas personas no fueron asignadas y cuántas tuvieron una asignación.

```
DEFINE PERSONA.TOTEVA
AS PERSONA.EVASEXO + PERSONA.EVAEDAD + ...
TYPE INTEGER
```

- d) La recodificación de la variable temporal anterior (la sumatoria), transformando los valores en 0 y 1 proporciona una estadística directa para el número de personas con y sin asignación.

```
DEFINE PERSONA.RECEVA
AS RECODE PERSONA.TOTEVA
(0 = 0)
ELSE 1
TYPE INTEGER
RANGE 0 – 1
VALUETAGS
0 "Sin cambios en la persona"
1 "Con cambios en la persona"
```

Este tipo de verificación se puede realizar en cada entidad y realizar combinaciones que permitan identificar más cambios, entre entidades, por ejemplo, sumar las personas que tuvieron asignación en cada vivienda, y con eso produciendo la frecuencia de viviendas sin asignación en las personas, o con por lo menos una asignación en las personas, etc.; También se pueden hacer análisis más sofisticados (tipo Índice de Myers, Razón de Sexo, etc.) para conjuntos separados de personas sin y con asignación

V. Anexo: Plan de tabulados y programas

```
RUNDEF Nuevo
  SELECTION ALL

//----- Vivienda -----/
TABLE VIV
  AS FREQUENCY
  OF TIPOLOC, PARED, TECHO, CONDOCUP, PISO

//----- Hogar -----/
TABLE HOG_GRAL
  AS FREQUENCY
  OF RESHLUG, RESHDPTO, RESHMPIO

TABLE HOG_EMIG
  AS FREQUENCY
  OF EMIGHOG, EMIGHOM, EMIGMUJ

TABLE HOG_SERV
  AS FREQUENCY
  OF Facilidades

TABLE HOG_DISCA
  AS FREQUENCY
  OF Discapacidad

//----- Persona -----/
TABLE PER_BAS
  AS FREQUENCY
  OF PARENTES, SEXO, EDAD, VIVEMAD

TABLE PER_ETNIA
  AS FREQUENCY
  OF Etnia

TABLE PER_EDUC
  AS FREQUENCY
  OF Educacion

TABLE PER_TRAB
  AS FREQUENCY
  OF Trabajo

TABLE PER_FEC
  AS FREQUENCY
  OF Fecundidad

TABLE PER_MIGRA
```

Ejemplos para evaluar la consistencia: Evaluación de EDAD del hijo mayor del hogar (respecto al jefe)

```
//Variable temporal de la edad del jefe del hogar
DEFINE HOGAR.EDAD_JEF
AS EDAD
FOR PARENTES = 1

//Variable temporal para la edad del hijo mayor en el hogar
DEFINE HOGAR.EDAD_HM
AS MAXIMUM EDAD
FILTER PARENTES = 3
TYPE INTEGER
RANGE 0 - 120

//Se deben cruzan las variables temporales para comparar que no exista en el hogar un hijo
con mayor edad que el jefe del hogar.

TABLE T1 AS FREQUENCY OF HOGAR.EDAD_JEF BY HOGAR.EDAD_HM

//El resultado debe ser una tabla donde la edad del hijo mayor (columna) aumenta con la edad
del jefe (fila)
```

Ocupación de las viviendas

```
//Variable derivada para la reclasificación de la ocupación de las viviendas//

DEFINE VIVIENDA.DV04
AS RECODE CONDOCUP
(1 = 1)
(2 = 2)
(3 - 6 = 3)
RANGE 1:3
VARLABEL "Condición de ocupación del local de habitación"
VALUETAGS
1 "Ocupado- personas presentes"
2 "Ocupado- personas ausentes"
3 "Desocupada"
```

Número de hogares por vivienda

```
//Variable derivada para conocer la cantidad de hogares en el local de habitación//  
  
DEFINE VIVIENDA.TOTHOG  
AS COUNT HOGAR  
  
DEFINE VIVIENDA.HOGVIV  
AS SWITCH  
INCASE TIPOLOC>6  
ASSIGN 0  
INCASE CONDOCUP > 1  
ASSIGN 0  
ELSE VIVIENDA.TOTHOG  
RANGE 1:9  
VARLABEL "Cantidad de hogares en el local de habitación"  
NOTAPPLICABLE 0
```

SEXO DEL JEFE DEL HOGAR

```
//Contar jefes de hogar  
DEFINE HOGAR.JH  
AS COUNT PERSONA  
FOR PARENTES = 1  
  
//Contar jefes de hogar hombres  
DEFINE HOGAR.SXJH  
AS COUNT PERSONA  
FOR PARENTES = 1 AND SEXO = 1  
  
//Contar jefes de hogar mujer  
DEFINE HOGAR.SXJM  
AS COUNT PERSONA  
FOR PARENTES = 1 AND SEXO = 2  
  
//Variable derivada para conocer el sexo del jefe del hogar//  
DEFINE HOGAR.SXJ  
AS SWITCH  
INCASE HOGAR.SXJH > 0  
ASSIGN 1  
INCASE HOGAR.SXJM > 0  
ASSIGN 2  
ELSE 0  
RANGE 1:2  
VARLABEL "Sexo del jefe del hogar"  
VALUETAGS  
1 "Jefe Hombre"  
2 "Jefe Mujer"
```


Agrupación de edades

```
//Variable derivada para agrupar edades en tres grandes grupos//  
DEFINE PERSONA.GPR3  
AS RECODE EDAD  
(0 - 14 = 1)  
(15 - 64 = 2)  
(65 - HIGHEST = 3)  
TYPE INTEGER  
RANGE 1:3  
  
VARLABEL "Grupo grandes de edad"  
VALUETAGS  
1 "Menores a 14 años"  
2 "De 15 a 64 años "  
3 "Mayores a 65 años"
```

```
//Variable derivada para agrupar edades en cuatro grandes grupos//  
DEFINE PERSONA.GPR4  
AS RECODE EDAD  
(0 - 14 = 1)  
(15 - 64 = 2)  
(65 - 79 = 3)  
(80 - HIGHEST = 4)  
TYPE INTEGER  
RANGE 1:4  
  
VARLABEL "4 grupos de edad"  
VALUETAGS  
1 "Menores a 14 años"  
2 "De 15 a 64 años "  
3 "Mayores a 65 años"  
4 "Mayores a 80 años"
```

```
//Mayores a 10 años por sexo
DEFINE PERSONA.ED_SX_MAYOR10
AS SWITCH
INCASE EDAD >=10 AND SEXO = 1
ASSIGN 1
INCASE EDAD >=10 AND SEXO = 2
ASSIGN 2
ELSE 0
TYPE INTEGER
RANGE 0:2
VARLABEL "Mayores a 10 años y más por sexo"
VALUETAGS
1 "Hombres de 10 años o más"
2 "Mujeres de 10 años o más"
0 "No aplica"
```

Población económicamente activa e inactiva (PEA PEI)

```
//Variable derivada para conocer la PEA y la PEI//
DEFINE PERSONA.PEA
AS SWITCH
INCASE EDAD < 7
ASSIGN 0
INCASE SEMANTRA = 1
ASSIGN 1
INCASE SEMANTRA = 2 AND SEMANACT < 8
ASSIGN 1
INCASE SEMANTRA = 2 AND SEMANACT >7 AND SEMANACT < 12)
ASSIGN 2
ELSE 0
RANGE 1:2
NOTAPPLICABLE 0
VARLABEL "Tipo de actividad económica"
VALUETAGS
1 "PEA"
2 "PEI"
```

Ocupación de las personas

```
//Variable derivada para conocer la condición de ocupación de las personas//  
DEFINE PERSONA.OCUP  
AS SWITCH  
INCASE EDAD < 7  
ASSIGN 0  
INCASE SEMANTRA = 1  
ASSIGN 1  
INCASE SEMANTRA = 2 AND SEMANACT < 6  
ASSIGN 1  
INCASE SEMANTRA = 2 AND (SEMANACT = 6 OR SEMANACT = 7)  
ASSIGN 2  
ELSE 0  
RANGE 1:2  
NOTAPPLICABLE 0  
VARLABEL "Condición de actividad de las personas "  
VALUETAGS  
1 "Ocupado"  
2 "Desocupado"
```

Otros indicadores para evaluar consistencia

```
//promedio de Hogares por vivienda  
DEFINE DEPTO.NHOG  
AS COUNT HOGAR  
  
DEFINE DEPTO.NVIV  
AS COUNT VIVIENDA  
FILTER TIPOLOC <6  
  
DEFINE DEPTO.HOG_VIV  
AS DEPTO.NHOG / DEPTO.NVIV  
FOR DEPTO.NVIV <>0  
TYPE REAL
```

Promedio de personas por hogar

```
DEFINE DEPTO.NPERS  
AS COUNT PERSONA  
  
DEFINE DEPTO.NHOG1  
AS COUNT HOGAR  
  
DEFINE DEPTO.PERS_HOG  
AS DEPTO.NPERS / DEPTO.NHOG  
FOR CPV2002.NHOG <> 0  
TYPE REAL
```

Porcentaje de viviendas ocupadas con personas ausentes

```
DEFINE DEPTO.NUM2  
AS COUNT VIVIENDA  
FILTER CONDOCUP = 2  
  
DEFINE DEPTO.DENOM2  
AS COUNT VIVIENDA  
FILTER CONDOCUP <=3  
  
DEFINE DEPTO.PCOCUP_AUS  
AS (DEPTO.NUM2 / DEPTO.DENOM2) * 100  
FOR DEPTO.DENOM2 > 0  
TYPE REAL
```

Porcentaje de viviendas desocupadas

```
DEFINE VIVIENDA.OCUP
AS RECODE CONDOCUP
(1 - 3 = 1)
(4 - 6 = 2)
RANGE 1:2
VARLABEL "CONDICION OCUPACION"
VALUELABELS
1 "OCUPADA"
2 "DESOCUPADA"

DEFINE DEPTO.NUM1
AS COUNT VIVIENDA
FILTER CONDOCUP = 2

DEFINE DEPTO.DENOM1
AS COUNT VIVIENDA

DEFINE DEPTO.PCOCUP
AS (DEPTO.NUM1 / DEPTO.DENOM1) * 100
FOR DEPTO.DENOM1 > 0
TYPE REAL
```

Porcentaje de población Urbano/Rural

```
//TOTAL POB URBANO
DEFINE DEPTO.POBURB
AS COUNT PERSONA
FILTER AREA = 1

//TOTAL POB RURAL
DEFINE DEPTO.POBRUR
AS COUNT PERSONA
FILTER AREA = 2

//TOTAL PERSONAS
DEFINE DEPTO.POBTOT
AS COUNT PERSONA

//% POB URBANA
DEFINE DEPTO.PCPOBURB
AS (DEPTO.POBURB / DEPTO.POBTOT) * 100
FOR DEPTO.POBTOT > 0
TYPE REAL

//% POB RURAL
DEFINE DEPTO.PCPOBRUR
AS (DEPTO.POBRUR / DEPTO.POBTOT) * 100
FOR DEPTO.POBTOT > 0
TYPE REAL
```

Índice de masculinidad

```
DEFINE DEPTO.SEXMASC
AS COUNT PERSONA
FILTER SEXO = 1

DEFINE DEPTO.SEXFEM
AS COUNT PERSONA
FILTER SEXO = 2

DEFINE DEPTO.INDMASCU
AS DEPTO.SEXMASC / DEPTO.SEXFEM
FOR DEPTO.SEXMASC >0
TYPE REAL
```

Relación Niños-Mujeres

```
//NIÑOS, VERIFICAR FILTRO EDAD
DEFINE DEPTO.NINOS
AS COUNT PERSONA
FILTER EDAD <=12

//MUJERES, VERIFICAR FILTRO EDAD
DEFINE DEPTO.MUJER
AS COUNT PERSONA
FILTER SEXO = 2 AND EDAD > 15

DEFINE DEPTO.NI_MUJER
AS DEPTO.NINOS / DEPTO.MUJER
FOR DEPTO.MUJER > 0
TYPE REAL
```

Relación de dependencia

```
//MENORES DE 15 Y MAYORES A 65
DEFINE DEPTO.ED15Y65
AS COUNT PERSONA
FILTER (EDAD <=14) OR (EDAD >=65)

//DE 15 A 64
DEFINE DEPTO.ED15_64
AS COUNT PERSONA
FILTER (EDAD >=15 AND EDAD <=64)

//RELACION
DEFINE DEPTO.REL_DEP
AS DEPTO.ED15Y65 / DEPTO.ED15_64
FOR DEPTO.ED15_64 > 0
TYPE REAL
```

Proporciones de grupos de edad

```
//MENORES DE 15
DEFINE DEPTO.ED15
AS COUNT PERSONA
FILTER EDAD <=14

//DE 15 A 64
DEFINE DEPTO.ED1564
AS COUNT PERSONA
FILTER (EDAD >=15 AND EDAD <=64)

//MAYORES 65
DEFINE DEPTO.ED65
AS COUNT PERSONA
FILTER EDAD >=65

//POB TOTAL
DEFINE DEPTO.POBTOTED
AS COUNT PERSONA

//PROPORCION MENORES 15
DEFINE DEPTO.PROP15
AS DEPTO.ED15 / DEPTO.POBTOTED
FOR DEPTO.POBTOTED > 0
TYPE REAL

//PROPORCION DE 15 A 64
DEFINE DEPTO.PROP1564
AS DEPTO.ED1564 / DEPTO.POBTOTED
FOR DEPTO.POBTOTED > 0
TYPE REAL

//PROPORCION MAYORES 65
DEFINE DEPTO.PROP65
AS DEPTO.ED65 / DEPTO.POBTOTED
FOR DEPTO.POBTOTED > 0
TYPE REAL
```


Proporción de inmigrantes internaciones en los últimos 5 años

```
DEFINE DEPTO.POBINM  
AS COUNT PERSONA  
FILTER PERSONA.P53C >=30 AND LLEGPAS >=1998
```

```
DEFINE DEPTO.POBTOTINM  
AS COUNT PERSONA  
FILTER PERSONA.P53C >=30
```

```
DEFINE DEPTO.PCPOBINM  
AS (DEPTO.POBINM / DEPTO.POBTOTINM) * 100  
FOR DEPTO.POBTOTINM > 0  
TYPE REAL
```

Personas que se autoidentifican como indígena/afrodescendiente

```
DEFINE DEPTO.POBIND  
AS COUNT PERSONA  
FILTER INDIGENA = 1
```

```
DEFINE DEPTO.POBTOT1  
AS COUNT PERSONA
```

```
DEFINE DEPTO.PCPOBIND  
AS (DEPTO.POBIND / DEPTO.POBTOT1) * 100  
FOR DEPTO.POBTOT1 > 0  
TYPE REAL
```

Población por nivel educativo

```
//1RA PARTE: CONTAR POBLACIÓN POR NIVEL EDUCATIVO
```

```
//POBLACION NO NIVEL EDUCATIVO
```

```
DEFINE DEPTO.POBNOEDUC
```

```
AS COUNT PERSONA
```

```
FILTER PERSONA.P13A = 1
```

```
//POBLACION NIVEL PREPRIMARIA
```

```
DEFINE DEPTO.POBPREP
```

```
AS COUNT PERSONA
```

```
FILTER PERSONA.P13A = 2
```

```
//POBLACION EDUCACIÓN PRIMARIA
```

```
DEFINE DEPTO.POBPRIM
```

```
AS COUNT PERSONA
```

```
FILTER PERSONA.P13A = 3
```

```
//POBLACION EDUCACIÓN MEDIA
```

```
DEFINE DEPTO.POBMEDIA
```

```
AS COUNT PERSONA
```

```
FILTER PERSONA.P13A = 4
```

```
//POBLACION EDUCACIÓN SUPERIOR
```

```
DEFINE DEPTO.POBSUPER
```

```
AS COUNT PERSONA
```

```
FILTER PERSONA.P13A = 5
```

```
//POBLACIÓN TOTAL PARA DENOMINADOR
```

```
DEFINE DEPTO.POBTOT2
```

```
AS COUNT PERSONA
```

//2DA PARTE: CALCULAR PORCENTAJES

// PORCENTAJE POBLACIÓN SIN NIVEL

DEFINE DEPTO.PCPOBNOEDUC

AS (DEPTO.POBNOEDUC / DEPTO.POBTOT2) * 100

FOR DEPTO.POBTOT2 > 0

TYPE REAL

// PORCENTAJE POBLACIÓN PREPRIMARIA

DEFINE DEPTO.PCPOBPREP

AS DEPTO.POBPREP / DEPTO.POBTOT2) * 100

FOR DEPTO.POBTOT2 > 0

TYPE REAL

// PORCENTAJE POBLACIÓN PRIMARIA

DEFINE DEPTO.PCPOBPRIM

AS (DEPTO.POBPRIM / DEPTO.POBTOT2) * 100

FOR DEPTO.POBTOT2 > 0

TYPE REAL

// PORCENTAJE POBLACIÓN MEDIA

DEFINE DEPTO.PCPOBMEDIA

AS (DEPTO.POBMEDIA / DEPTO.POBTOT2) * 100

FOR DEPTO.POBTOT2 > 0

TYPE REAL

// PORCENTAJE POBLACIÓN SUPERIOR

DEFINE DEPTO.PCPOBSUPER

AS (DEPTO.POBSUPER / DEPTO.POBTOT2) * 100

FOR DEPTO.POBTOT2 > 0

TYPE REAL