



10 de marzo de 2006

MEMO PARA Peter O. Way
Jefe, Centro de Programas Internacionales

A Través de: Kevin E. Deardorff *VED*
Asistente del Jefe del Centro de Programas Internacionales para
Asistencia Técnica y Capacitación

De: Glenn Ferri *Glenn Ferri*
Jefe, Oficina de Metodología y Desarrollo de Software

Motivo: Aprobación de Informe de Misión

El informe de misión de Glenn Ferri a Managua, Nicaragua, 6-18 de marzo de 2006, y en Washington del 20-24 de marzo de 2006, se adjunta para su aprobación.

Adjuntos

APROBADO: *Peter O. Way* FECHA: 14 MAR 06
PETER O. WAY

APROBADO: *Armando R. Levinson* FECHA: 14 MARZO 2006
ARMANDO R. LEVINSON (P.M.)

DISTRIBUCIÓN RECOMENDADA

cc:	N. Delgadillo	INEC/Nicaragua
	C. Goffin	INEC/Nicaragua
	M. Beteta	INEC/Nicaragua
	H. Larai	IDB/Nicaragua
	R. Singh	USAID/GH/PRH/PEC

Jefe Interino de la División de Población
K. Deardorff
M. Therrien
D. Lopez-Meisel
A. Levinson
Archivos



NICARAGUA INFORME DE MISION

Preparado por: Glenn Ferri

Fecha del Informe: 10 de marzo de 2006

Lugar Visitado: Managua, Nicaragua

Presentado a: Christine Goffin, INEC

Período Covered: 6 al 24 de febrero, 2006

Propósito de la Misión y Resumen

A pedido del Banco Inter-Americano de Desarrollo y del Instituto Nacional de Estadística y Censos (INEC) de Nicaragua, Glenn Ferri viajó a Managua, Nicaragua, para evaluar el sistema de procesamiento de datos preparado para el VIII Censo de Población y IV de Vivienda (CPV). El CPV tuvo lugar durante el período del 28 de mayo al 11 de junio, 2005.

A su llegada, Ferri se reunió con Santiago Mejía, Jefe del Departamento de Análisis y Diseño de Procesamiento de Datos, para discutir el propósito de la visita. Mejía le presentó a Ferri los otros miembros del personal de procesamiento de datos. También le proporcionó a Ferri todos los programas con bastante detalle, así como las especificaciones de crítica y tabulación, y otros documentos relevantes al sistema de procesamiento de datos censales. Ferri revisó todos estos materiales, se reunió con miembros del INEC y escribió varios programas *ad hoc* para evaluar el sistema.

Luego de pasar dos semanas en Managua, Ferri regresó a Washington y pasó una semana más analizando todo el sistema de procesamiento de datos y presentando sus hallazgos preliminares a otros colegas para la revisión final. Ferri preparó un informe que resume todos los hallazgos. Dicho informe se encuentra en el Anexo B. La lista de personas contactadas puede verse en el Anexo A.

Adjuntos

cc:

N. Delgadillo	INEC/Director
C. Goffin	INEC/Coordinadora del Censo
M. Beteta	INEC/Director Técnico del Censo
H. Larai	BID/Managua

ANEXO A

Personas Contactadas

Instituto Nacional de Estadística y Censos (INEC)

Domicilio: Frente al Hospital Lenín Fonseca
Managua, Nicaragua

Teléfono: 505-266-6564
Fax: 505-266-2031

Nestor Delgadillo, Director del INEC
Christine Goffin, Coordinadora Censos Nacionales Coordinator (cgoffin@inec.gob.ni)
Santiago Mejia Solis, Jefe del Depto. De Procesamiento de Datos
Margel Beteta, Director Técnico del Censo (margel@inec.gob.ni)
Jimmy Rosales, Estadístico
Domingo Primante, Consultor en estadística
Odette Robleto, Programador
Javier Argeñal, Programador
Mario Neón Solis, Programador

Banco InterAmericano de Desarrollo (BID)

Domicilio: Km. 4 ½ Carretera a Masaya (next door to the Princess Hotel)
Managua, Nicaragua

Teléfono: 505-267-0831
Fax: 505-267-3469

Hazel Larai, Especialista del Sector Social (hazelp@iadb.org)
Karla Patricia Arriola Mercado, Muestrista (karriola@inec.gob.ni)

ANEXO B

**Evaluación del Sistema de Procesamiento de Datos del
Censo de Población y Vivienda de 2005**

República de Nicaragua

Marzo 2006

Glenn Ferri

International Programs Center

U. S. Census Bureau

Indice

Resumen Ejecutivo	i
1. Antecedentes	1
3. Preparación para el Procesamiento Electrónico	1
3.1. Sistema de Control Operacional	1
3.2. Depuración Manual	2
3.3. Resumen de Observaciones	3
4. Entrada de Datos y Verificación	3
4.1. Digitación de Datos	3
4.2. Aplicación para la Entrada de Datos	3
4.3. Estadísticas del Digitador	4
4.4. Resumen de Observaciones	5
5. Depuración e Imputación	5
5.1. Consolidación de Datos para la Depuración	5
5.2. Consistencia Estructural	5
5.3. Consistencia e Imputación	7
5.4. Resumen de Observaciones y Recomendaciones	8
6. Tabulación	9
6.1. Plan de Tabulación	9
6.2. Programas de Tabulación	9
6.3. Resumen de Observaciones y Recomendaciones	10
7. Productos para la Difusión de Datos	10

Resumen Ejecutivo

El propósito de este informe es documentar la evaluación del sistema de procesamiento de datos preparado por el Instituto Nacional de Estadísticas y Censos (INEC) de Nicaragua para procesar los datos del VIII Censo de Población y IV de Vivienda (CPV). El CPV se llevó a cabo durante el período del 28 de mayo al 11 de junio de 2005.

Ferri viajó a Managua, Nicaragua, por dos semanas para trabajar con el equipo del INEC. Le proporcionaron a Ferri todos los programas bien detallados, las especificaciones de crítica y de tabulación y otros documentos relevantes sobre el sistema de procesamiento de datos del censo. Ferri revisó todos estos materiales, se reunió con miembros del INEC y escribió varios programas ad hoc para evaluar el sistema. Luego, Ferri regresó a Washington y pasó una semana más analizando el sistema de procesamiento de datos y haciendo que sus colegas en Washington revisaran en detalle los hallazgos preliminares. Este informe es el resultado de este proceso. A continuación se ofrece un resumen de observaciones y recomendaciones claves.

- El personal de procesamiento de datos es pequeño pero todos están bien preparados. Se les debe felicitar por el excelente trabajo que han hecho hasta el momento para el procesamiento de los datos censales. Han mostrado tener conocimiento de las mejores prácticas en procesamiento de datos y una gran habilidad y dedicación al producir los programas requeridos. El personal de procesamiento de datos representa un componente clave de lo que seguramente será un exitoso censo en Nicaragua.
- Debido en parte al éxito del sistema de procesamiento de datos, el INEC parece poder cumplir con el calendario de publicación de datos censales en abril de 2006, como se tenía originalmente planificado.
- El sistema de control operacional fue muy efectivo ya que produjo informes sobre el estado de cada operación pudiéndose, de esa manera, rastrear los cuestionarios a través de todas las fases de procesamiento.
- La operación de crítica manual se llevó a cabo a tiempo, pero así y todo quedaron algunos códigos no válidos. El equipo del INEC está usando buenos procedimientos para corregir estos códigos.
- El grupo de procesamiento de datos del INEC utilizó las herramientas correctas para procesar el censo, especialmente el CSPro (*Census and Survey Processing System*).
- La operación de entrada y verificación de datos fue bien diseñada y ejecutada, usando las mejores prácticas estándar.
- El INEC aplicó un buen mecanismo de control de calidad en las operaciones, especialmente en el uso de estadísticas de los digitadores de datos.
- La aplicación para la entrada de datos fue bien diseñada y programada. Ferri quedó impresionado con el personal del INEC ya que hicieron uso completo de las funciones avanzadas de CSPro y tenían un conocimiento profundo del sistema.
- La alta calidad de la aplicación para la entrada de datos y los Buenos procedimientos operacionales para la entrada de datos dieron como resultado una alta calidad en los datos digitados.
- El INEC está utilizando correctamente la metodología de crítica e imputación.

- Los especialistas en la materia y los programadores trabajan en conjunción; hay buenas especificaciones de crítica y los programas están bien escritos.
- Los programadores del INEC tienen un muy buen conocimiento de las técnicas de programación de crítica e imputación, así como de la programación de aplicaciones de lote de CSPro.
- Las especificaciones estructurales son sólidas y la dificultad de lidiar con datos a nivel de vivienda y a nivel de hogar ha sido satisfactoriamente resuelta.
- El programa de crítica e imputación todavía no está terminado, como lo demuestra los errores obtenidos en la segunda corrida de los archivos de Managua y Nueva Segovia. El INEC debe resolver estos problemas.
- El INEC debe rutinariamente correr el programa de crítica e imputación dos veces para cada archivo de datos.
- El INEC debe crear un diccionario de datos definitivo para probar el programa de crítica, para definir los valores válidos de cada variable después de la imputación, prestando atención especial a aquellas variables que puedan o no aceptar valores en blanco. Se debe correr el módulo de frecuencias de CSPro (*Frequencies*) incluyendo los valores no válidos.
- El INEC debe correr frecuencias para todas las variables antes y después de correr el programa de crítica e imputación para verificar doblemente que las imputaciones no hayan sesgado los datos.
- Los especialistas en la materia y los programadores cooperan mutuamente muy bien; las especificaciones del plan de tabulación son buenas y los programas están bien escritos.
- El personal de programación del INEC tiene un muy buen conocimiento de las técnicas de programación de planes de tabulación usando el módulo CENTS de IMPS.
- El personal de programación del INEC debe continuar aprendiendo el uso de CSPro para llevar a cabo planes de tabulación. Deben correr los programas de tabulación en CENTS y comparar los resultados con los obtenidos en CSPro.

1. Antecedentes

El propósito de este informe técnico es evaluar el sistema de procesamiento de datos preparado por el personal del Instituto Nacional de Estadísticas y Censos (INEC) de Nicaragua para poder llevar a cabo el procesamiento de los datos censales del VIII Censo de Población y IV de Vivienda (CPV) de 2005. Se había planificado originalmente levantar el censo durante el período del 25 de abril al 8 de mayo de 2005. Sin embargo, debido a circunstancias imprevistas, se pospusieron las fechas y el censo finalmente se levantó entre el 28 de mayo y el 11 de junio de 2005.

En noviembre/diciembre del 2005, Armando Levinson, del Buró de Censos de los Estados Unidos, llevó a cabo una evaluación más profunda del censo de población. Su informe titulado "Documentación de las Actividades Más Importantes del Censo de Población y Vivienda de 2005" fue presentado en diciembre de 2005. La evaluación de Ferri es más específica ya que examina el sistema de procesamiento de datos en detalle.

Al momento de la visita de Ferri, todo el trabajo de campo ya se había completado. Los datos ya habían pasado por la crítica manual y ya se habían capturado por medio de digitación. También existía una primera versión de los programas de crítica y de tabulación y ya se habían corrido dichos programas con datos censales. Debido a la naturaleza iterativa del procesamiento de datos censales, todavía se estaban refinando los programas de crítica y tabulación a medida que el INEC revisaba los resultados. El INEC tiene planeado publicar los resultados del censo en abril de 2005.

Como al momento de escribir este informe todavía continuaba el proceso de crítica y tabulación, Ferri puede ofrecer observaciones y recomendaciones en estas dos áreas. Para las fases anteriores, que ya se han completado, Ferri sólo puede ofrecer observaciones.

2. Recursos Humanos para el Procesamiento de Datos

El INEC tiene tres programadores asignados a trabajar en el sistema de procesamiento de los datos censales. El programador/analista principal tiene una gran habilidad y suma experiencia en el procesamiento de datos ya que él estuvo involucrado en el censo de 1995. El personal del INEC posee un conocimiento completo del programa IMPS (*Integrated Microcomputer Processing System*), ISSA (*Integrated System for Survey Analysis*), y CSPro (*Census and Survey Processing System*). IMPS e ISSA son dos programas que utilizan el sistema operativo DOS y ambos han sido reemplazados por el programa CSPro que utiliza la plataforma Windows. Los otros programadores tienen también conocimiento y experiencia en CSPro. Uno de los programadores ha aprendido el módulo de tabulación de IMPS que se utiliza para la tabulación de datos censales. El personal también tiene conocimiento de otros paquetes de procesamiento de datos como SPSS, SQL y FoxPro.

El personal es pequeño pero todos están bien preparados. Se les debe felicitar por el excelente trabajo que han hecho hasta el momento para el procesamiento de los datos censales. Han mostrado tener conocimiento de las mejores prácticas en procesamiento de datos y una gran habilidad y dedicación al producir los programas requeridos. El personal de procesamiento de datos representa un componente clave de lo que seguramente será un exitoso censo en Nicaragua. Ferri recibió toda la cooperación posible al recibir los programas y otros archivos pertinentes para llevar a cabo su evaluación.

3. Preparación para el Procesamiento Electrónico

3.1. Sistema de Control Operacional

Un sistema de control operacional efectivo es un elemento crítico para el éxito del censo. Dicho sistema de control permite asegurar que la cobertura sea completa, permite monitorear el flujo de cuestionarios a través de todas las fases de procesamiento manual y electrónico y produce informes actualizados del estado de la operación completa.

Nicaragua está dividida administrativamente en 15 Departamentos más dos regiones autónomas: *Región Autónoma del Atlántico Norte* (RAAN) y *Región Autónoma del Atlántico Sur* (RAAS). Los Departamentos se dividen en 153 Municipios. A su vez, los Municipios se dividen en 8,603 segmentos censales (áreas de empadronamiento). El segmento es la unidad básica del control operacional.

Los cuestionarios regresan del campo a la oficina central por segmento o lote. Cada lote incluye una Hoja de Listado del Área de Empadronamiento que se llama *Hoja de Recorrido del Segmento* u *Hoja de Ruta del Segmento*. Este formulario contiene información sobre cada hogar en el segmento, incluyendo el nombre del jefe del hogar, el resultado de la visita y, muy importante, el número de hombres y mujeres en el hogar.

Los cuestionarios pasan luego a los *ordenadores*, cuyo trabajo consiste en comparar la cartografía con las hojas de ruta y re-numerar todas las viviendas para que sigan un orden lógico. Este paso es necesario ya que había casos de viviendas que se encontraron en el campo pero que no aparecían en los mapas de las áreas de empadronamiento.

Después de esta fase, los cuestionarios pasan a mano de los codificadores que llevan a cabo la crítica manual descrita en la sección siguiente y codifican algunas variables, como ocupación y rama de actividad. Finalmente, los cuestionarios pasan a mano de los supervisores de digitación para la subsiguiente entrada de datos.

Ferri observó que el sistema de control operacional era bien efectivo. El personal del INEC pudo muy rápidamente producir informes que indicaban el estado de cada operación censal. Más significativamente aún, el INEC pudo rastrear con certeza y exactitud cada segmento a través de cada fase del procesamiento para determinar el alcance preciso del daño causado por el desafortunado robo de algunos cuestionarios censales.

3.2. Depuración Manual

Algunos campos en el cuestionario requieren depuración manual. Esto significa que cada respuesta numérica en cada cuestionario debe ser examinada y seguidamente codificada antes de pasar al proceso de digitación. Estos campos incluyen:

- País (Sección IV, Pregunta 4)
- Departamento y Municipio (Sección VII, Preguntas 9, 11 y 26)
- Estudio universitario (Sección VII, Pregunta 18)
- Ocupación (Sección VII, Pregunta 28)
- Rama de Actividad (Sección VII, Pregunta 30)

La Sección VI-A contiene una lista de las personas en el hogar, por nombre y sexo. Esta información crucial asegura durante el procesamiento electrónico que las cifras de población por género sean lo más exactas posibles. El personal de depuración manual verificó el código de la variable sexo en cada cuestionario y cuando el mismo se encontraba ausente se llenaba teniendo en cuenta el nombre de la persona u otras características en el cuestionario.

Durante el operativo de campo hubo casos en que se encontraron unidades de vivienda que no aparecían en los mapas de las áreas de empadronamiento. En estos casos, se asignaron códigos temporales que fueron seguidamente reordenados, para el segmento completo, en la oficina.

La depuración manual fue llevada a cabo por 144 codificadores que trabajaron en 2 turnos bajo la supervisión de 6 jefes de codificación y 24 supervisores de codificación. La depuración manual terminó en diciembre de 2005.

Ferri observó que el INEC está haciendo un esfuerzo metódico para investigar códigos no válidos en campos que fueron manualmente depurados. Los programadores escribieron programas para listar todos

los cuestionarios con códigos no válidos. Estos cuestionarios fueron seguidamente recuperados para corregir códigos incorrectos en las listas. Estas correcciones fueron entradas en otro programa CSPro que corregía el error en la base de datos.

3.3. Resumen de Observaciones

- El sistema de control operacional fue muy efectivo ya que produjo informes sobre el estado de cada operación pudiéndose, de esa manera, rastrear los cuestionarios a través de todas las fases de procesamiento.
- La operación de crítica manual se llevó a cabo a tiempo, pero así y todo quedaron algunos códigos no válidos. El equipo del INEC está usando buenos procedimientos para corregir estos códigos.

4. Entrada de Datos y Verificación

4.1. Digitación de Datos

Los datos censales fueron digitados a computadoras personales por un equipo de 60 digitadores que trabajaron en 2 turnos, con un supervisor por cada 12 digitadores. Los datos fueron digitados por segmento y cada segmento se registraba en un archivo de datos separado. La aplicación de digitación se escribió en CSPro.

El INEC utilizó el modulo de CSPro que permite la verificación dependiente (doble entrada de datos) para minimizar errores de digitación. Esto significa que después de haber digitado un segmento, otro digitador (verificador) entra los mismos datos por segunda vez. A medida de que el verificador digita los datos, CSPro compara el valor de cada campo con el valor originalmente digitado y advierte al verificador si ambos valores no concuerdan. Luego, el verificador adjudica el valor correspondiente. El INEC eligió el segmento para verificar y verificó todos los casos dentro de esos segmentos. El INEC verificó el 60% de los segmentos de Managua y el 30% del resto del país. Esta es una estrategia razonable de verificación.

Como el trabajo del verificador es más importante (él decide cuando hay discrepancia), el INEC eligió como verificadores a los mejores digitadores de datos durante la primera entrada de datos. Esto es lo que se recomienda en la práctica.

4.2. Aplicación para la Entrada de Datos

Ferri examinó la aplicación para la entrada de datos que se diseñó en CSPro para la fase de digitación de datos y verificación. Observó detenidamente el diccionario de datos, los formularios, los textos de pantalla y la lógica subyacente en el programa. Ferri digitó varios cuestionarios ficticios para verificar el comportamiento del programa bajo varias condiciones y rutas de datos.

El enfoque apropiado al diseñar una aplicación para la entrada de datos es permitirle al operador transcribir del cuestionario a la computadora, de la manera más rápida posible, los datos tal y como aparecen en el cuestionario de papel. Existe un mínimo de depuración durante la captura para poder maximizar la productividad. Toda inconsistencia en los datos se corrige electrónicamente en una fase posterior.

La aplicación para la entrada de datos del INEC está muy bien diseñada e implementada. La aplicación es de tipo “controlada por el operador,” que, como el mismo nombre lo sugiere, permite que el operador tenga libertad sustancial para moverse rápidamente entre los campos. Se han implementado correctamente controles de rangos en base a valores almacenados en el diccionario de datos.

La aplicación contiene algunos controles de crítica para los códigos geográficos entrados. Este control es muy bueno porque es extremadamente importante que los códigos geográficos estén correctos ya que éstos definen, en gran parte, la estructura de los datos y códigos incorrectos pueden ocasionar dificultades.

La aplicación también contiene un control de crítica para la lista de personas en el hogar y sus respectivos sexos en la Sección IV-A. Este también es un buen control ya que estos campos se usan más tarde en la consistencia de datos para verificar recuentos de personas por sexo. Estos recuentos son quizás las salidas más fundamentales de un censo y, por lo tanto, vale la pena tomar medidas extras para asegurar la exactitud de estos datos.

La aplicación de estos controles producen mensajes de error para el operador por medio de las funciones de CSPro “errmsg” y “accept.” Estos mensajes tienen como propósito alertar al operador para evitar que el mismo entre datos inconsistentes. Esto también es correcto ya que el objetivo no es forzar la entrada de datos limpios, sino transcribir los datos que se encuentran en el cuestionario.

Hay muy pocos saltos automáticos programados en la aplicación. Esto es importante porque la entrada de datos censales es una operación denominada de “cabeza abajo”; es decir, el operador generalmente mira el cuestionario a medida que va digitando y utiliza ciertos ritmos para digitar rápidamente. Los saltos automáticos pueden romper el ritmo y causar errores de digitación.

Sin embargo, hay circunstancias en que el operador elige un salto hacia otra sección. En estos casos, la aplicación correctamente usa “saltos del operador.” Los programadores del INEC mostraron bastante ingenio en el uso del procedimiento “OnKey” de CSPro para procesar la tecla “+” (salto del operador) en ciertas situaciones en que no funciona automáticamente. El uso de esta función de CSPro bien avanzada muestra el conocimiento profundo de CSPro que tiene el personal del INEC. Este conocimiento también se evidencia por el uso correcto de procedimientos lógicos, de campos persistentes y protegidos, y de elementos del lenguaje CSPro bien sofisticados.

La aplicación también hace buen uso del diseño de pantalla, de casillas y del color. Las pantallas de entrada de datos se han diseñado para que generalmente, pero no estrictamente, se parezcan a los cuestionarios impresos. Hay un poco de desplazamiento dentro de las pantallas, lo que no es deseable. La aplicación utiliza de manera muy interesante la función “CAPI” de CSPro. Se utiliza generalmente esta función para exhibir la pregunta que usa el empadronador en la parte superior de la pantalla en aplicaciones CAPI (Computer Assisted Personal Interviewing, o sea Entrevista Personal con Ayuda de Computadora). Aunque esta aplicación en particular no es CAPI, usa esta función para mostrar los títulos de las secciones del cuestionario. Esta es, en parte, la razón por la cual hay desplazamiento en algunos formularios.

4.3. Estadísticas del Digitador

CSPro mantiene estadísticas del digitador que miden la velocidad (teclas por hora y cuestionarios a través del tiempo) y exactitud (errores de digitación y campos cambiados por el verificador) en el archivo .LOG. Estos archivos son muy útiles al momento de evaluar el rendimiento de un digitador y de medir la calidad de la operación de entrada de datos. Durante la operación de entrada de datos, el INEC cargó todos los archivos de tipo .LOG en una base de datos y escribió una aplicación en FoxPro para producir informes semanales para cada digitador. Estos informes también proporcionan indicadores por cuartiles para todos los operadores y los valores extremos se marcan en rojo. Los supervisores de entrada de datos revisan estos informes regularmente.

Ferri examinó el informe semanal del 5-12 de noviembre de 2005. El informe está bien estructurado y contiene la información correcta para los supervisores. Durante esa semana, los operadores promediaron 5,800 pulsaciones por hora, que es una velocidad aceptable. Esto se traduce en alrededor de 18.5 cuestionarios por hora. Los verificadores cambiaron alrededor de un 2.2% de campos que fueron originalmente digitados. Este es un nivel bajo e indica una alta calidad en entrada de datos.

4.4. Resumen de Observaciones

- La operación de entrada y verificación de datos fue bien diseñada y ejecutada, usando las mejores practicas estándar.
- El INEC aplicó un buen mecanismo de control de calidad en las operaciones, especialmente en el uso de estadísticas de los digitadores de datos.
- La aplicación para la entrada de datos fue bien diseñada y programada. Ferri quedó impresionado con el personal del INEC ya que hicieron uso completo de las funciones avanzadas de CSPro y tenían un conocimiento profundo del sistema.
- La alta calidad de la aplicación para la entrada de datos y los buenos procedimientos operacionales para la entrada de datos dieron como resultado una alta calidad en los datos digitados.

5. Depuración e Imputación

5.1. Consolidación de Datos para la Depuración

Los datos fueron digitados y verificados por segmento. Se creó un archivo de datos para cada segmento a medida que se digitaba, y luego se verificaba (se digitaba por segunda vez). Los nombres de los archivos de datos contienen los códigos de departamento, municipio, área de supervisión y segmento. Esta es una buena técnica de nomenclatura que asegura un rastreo fácil.

El INEC usa la utilidad de CSPro llamada CSConcat para concatenar los archivos de segmentos en un solo archivo a nivel de Departamento. Luego, se usó la utilidad de CSPro llamada CSSort para ordenar los datos por código geográfico.

Para examinar los pasos restantes en el sistema de procesamiento de datos, Ferri trabajó con los archivos de dos Departamentos: Managua y Nueva Segovia. Managua, que incluye la ciudad capital, es mayormente urbano (86%), mientras que Nueva Segovia, frontera con Honduras, es mayormente rural (64%).

Ferri descubrió que habían 445 registros en blanco (códigos de tipo de registro presentes, pero sin identificación) en el archivo de datos de Nueva Segovia (registros del 8383 al 8827). Esto indica que hay un problema en la concatenación y en el ordenamiento que el INEC debe examinar cuidadosamente. Ferri eliminó los registros en blanco del archivo de datos antes de proceder con los pasos restantes.

5.2. Consistencia Estructural

La depuración de datos comienza con la verificación de una estructura correcta para cada cuestionario. Este involucra la verificación de códigos geográficos, la presencia de registros requeridos y la búsqueda de registros erróneos. El INEC ha logrado esto en dos pasos.

El primer paso es identificar y corregir códigos geográficos en blanco. El INEC usa una aplicación de lote sencilla de CSPro ("LlenarBlancosID.bch"). Ferri corrió esta aplicación usando los archivos de Managua y Nueva Segovia. Sólo se encontraron 4 códigos geográficos en blanco en Nueva Segovia y 41 en Managua. Esto está en el orden de .01% de casos en ambos archivos de datos, un número extremadamente pequeño.

El segundo paso fue encargarse de los tipos de registros. En el caso de este censo, esta parte de la depuración estructural se complica por el hecho de que los datos censales se registran en registros de viviendas y en registros de hogares. En general, una unidad de vivienda es una estructura física, mientras que un hogar es una familia (o un individuo) dentro de una unidad de vivienda. La mayoría de

las unidades de vivienda consisten en un solo hogar, pero hay casos en que una unidad de vivienda contiene más de un hogar. Esta es la primera vez que un censo en Nicaragua captura la información de esta manera.

En el cuestionario impreso se captura un solo hogar. En el caso de hogares múltiples en una unidad de vivienda, los cuestionarios múltiples se vinculan. Se registra la unidad de vivienda en el primer cuestionario, pero no en los subsiguientes.

Esta estructura jerárquica es común en censos de población y vivienda que capturan datos tanto para la vivienda como para el hogar y presenta un problema en la organización de datos para propósitos de digitación y su subsiguiente procesamiento. Por lo tanto, la resolución correcta de este problema es extremadamente importante para evitar la introducción de errores más tarde en el procesamiento.

El INEC ha resuelto este problema de manera muy inteligente usando las capacidades plenas de CSPro. El diccionario de datos para la entrada de datos correctamente define un caso como un hogar (un cuestionario impreso). Cuando una vivienda tiene más de un hogar, éstos se entran como casos separados de tal manera que el segundo hogar y los subsiguientes (si hay) tienen datos de vivienda en blanco. Luego de haber digitado los datos, se reformatean usando un diccionario de datos algo diferente en el cual un caso se define como una unidad de vivienda. Cuando una unidad de vivienda tiene hogares múltiples, el caso correspondiente (de acuerdo al nuevo diccionario) tendrá múltiples registros de vivienda, el primero con datos válidos y los otros sin datos válidos. Una aplicación de lote de CSPro ("EliminarViv.bch") identifica y elimina los registros de vivienda extras, así como todos los otros registros en blanco de otro tipo que puedan haberse introducido en la base de datos que fue digitada.

Ferri revisó estos diccionarios de CSPro y el programa de reformateo (EliminarViv.bch). Ferri corrigió un pequeño error en el programa que causaba un número pequeño de errores de subíndices al momento de su ejecución. También agregó denominadores a las estadísticas de salida para entender mejor el alcance de las correcciones hechas por la aplicación. Luego, Ferri corrió el programa usando los datos de Managua y Nueva Segovia. El programa arrojó los siguientes resultados:

NUEVA SEGOVIA:

User unnumbered messages: (mensajes no numerados del usuario)

Line	Freq	Pct.	Message text	Denom
51	1194	2.5	BORRAR REGISTRO VIVIENDA i=%d EN BLANCO...	47510
70	55	0.1	BORRAR SEGUNDO REGISTRO VIVIENDA i=%d...	47510
104	888	17.8	BORRAR REGISTRO EMIGRACION i=%d EN BLANCO...	4976
120	28	3.0	BORRAR REGISTRO DEFUNCION i=%d EN BLANCO...	949
134	2496	1.2	BORRAR REGISTRO PERSONA...	202724
173	1	0.0	BORRAR REGISTRO HOGAR i=%d V01=%d V02=%d Perso...	40858

CSPRO Executor Normal End

MANAGUA:

User unnumbered messages:

Line	Freq	Pct.	Message text	Denom
51	26377	8.9	BORRAR REGISTRO VIVIENDA i=%d EN BLANCO...	297805
70	3514	1.2	BORRAR SEGUNDO REGISTRO VIVIENDA i=%d...	297805
104	4913	10.2	BORRAR REGISTRO EMIGRACION i=%d EN BLANCO...	48349
120	665	7.8	BORRAR REGISTRO DEFUNCION i=%d EN BLANCO...	8552
134	62832	4.9	BORRAR REGISTRO PERSONA...	1284964
173	10	0.0	BORRAR REGISTRO HOGAR i=%d V01=%d V02=%d Perso...	270516

CSPRO Executor Normal End

Se espera la eliminación de registros de vivienda. El número de registros de hogar en blanco que se eliminó es insignificante, lo que es muy bueno. El número de personas en blanco y de registros de

mueres indica una digitación incorrecta. El alto número de registros de emigración eliminado es preocupante e indica una falta de atención por parte de los digitadores.

El INEC ha tratado de ordenar los datos de tal manera que se ajusten a un diccionario de datos jerárquico de dos niveles para la depuración y la tabulación, pero surgieron problemas que obligaron a abandonar este enfoque. Ferri identificó un descuido en el ordenamiento (falta de inclusión de la variable de hogar en el registro de vivienda en el nuevo diccionario). Una vez que se corrija este problema, se podrá quizás procesar a dos niveles. Sin embargo, un diccionario jerárquico de dos niveles no es necesario, y el uso de relaciones en el diccionario de un nivel (no jerárquico) resuelve igualmente este problema. Los programadores del INEC entienden las relaciones de CSPro y las están usando correctamente.

De cualquier manera, Ferri verificó que los archivos de datos reformateados sean apropiados para el ulterior proceso de depuración y tabulación.

5.3. Consistencia e Imputación

Una vez que los datos hayan sido correctamente estructurados, el siguiente paso es asegurar que los datos dentro de un hogar sean internamente consistentes, de acuerdo a reglas predeterminadas. Por ejemplo, una persona de tres años de edad no puede haber tenido hijos. Si los datos no pasan las reglas de consistencia, se debe cambiar uno o más datos (hay que imputar), nuevamente, de acuerdo a reglas predeterminadas implementadas en un programa de computación. Esto es necesario para que las tabulaciones y otros análisis de datos tengan sentido.

La tarea de asegurar consistencia interna es probablemente la más difícil dentro del sistema completo de procesamiento de datos. Las reglas para verificar la consistencia y especialmente para corregir los datos son extremadamente complejas. Las correcciones se llevan a cabo examinando otras variables en el hogar y, a veces, importando valores de datos de hogares vecinos. Esto último se lleva a cabo por medio de una técnica conocida como “hot-decking” o imputación dinámica.

La práctica estándar dicta que los expertos en la materia preparen las reglas de crítica e imputación. Estas deben ser completas, comprensivas y estar completamente documentadas. Estas reglas se pasan a los programadores para que puedan escribir los programas y luego correrlos. Esto es, en general, un proceso iterativo que resulta en un conjunto de especificaciones de crítica que se retroalimentan. Los programadores identifican lagunas y/o ambigüedades en las especificaciones y las mismas se regresan a los expertos en la materia quienes seguidamente modifican las especificaciones. Una vez que se corren los programas con datos reales, surgen situaciones en los datos que no fueron contempladas en las especificaciones y, por lo tanto, hay que seguir actualizando las especificaciones y los programas. Es muy importante que los expertos en la materia y los programadores trabajen estrechamente a través de todo este proceso.

El INEC todavía está por finalizar las especificaciones de crítica y los programas. Ferri se reunió con los expertos en la materia y examinó las especificaciones de crítica. Las mismas están bien formuladas y bien documentadas. Parece haber una buena comunicación entre estos dos grupos y los cambios en las especificaciones se están documentando correctamente. Todo parece indicar que se están utilizando las mejores prácticas.

Ferri examinó las aplicaciones de lote de CSPro (“CorreccionData.bch”) desarrolladas por el INEC para la crítica y la imputación. Nuevamente, Ferri quedó impresionado con el profundo conocimiento de CSPro que tiene el personal del INEC. El programa hace uso extenso de la técnica de imputación dinámica para imputar valores a partir de hogares cercanos. Esto requiere la manipulación de varias matrices y se implementan correctamente por medio del uso del sistema de archivo de búsqueda externa de CSPro. El programa utiliza técnicas apropiadas para producir tasas de error y frecuencias de imputación.

Ferri corrió el programa usando como insumo los archivos de Managua y de Nueva Segovia. Ferri

observó primeramente las tasas de error e imputación, que se muestran en las Tablas A01 y A02 del Adjunto A de este anexo. Las tasas son extraordinariamente bajas—por debajo del 5% para todos los casos—lo que indica un empadronamiento de alta calidad, así como de depuración manual y digitación. Las variables claves, Parentesco, Edad y Sexo, muestran particularmente bajos niveles de imputación (1% o menos para Managua), aunque las rutinas de imputación no se han completado, como se puede ver en las Tablas A05 a A08 del Adjunto A. Ferri confirmó estos resultados escribiendo una aplicación de lote independiente y verificando que se obtuvieran los mismos resultados.

Una prueba importante para monitorear el programa de crítica e imputación es correrlo dos veces. Ferri llevó a cabo esta prueba usando los archivos de Managua y Nueva Segovia. Tomó el archivo de salida de la primera corrida y lo utilizó como archivo de entrada para la segunda corrida. Si el programa de crítica e imputación funciona correctamente, no debería haber ningún mensaje de error proveniente de la segunda corrida ya que los datos debieron quedar depurados durante la primera corrida. Los resultados provenientes de la segunda corrida se encuentran en las Tablas A03 y A04 del Adjunto A. En general, los datos quedaron limpios después de la primera corrida. Sin embargo, hubo un número pequeño de imputaciones que se hicieron en la segunda corrida en ambos archivos. Esto indica una especificación no válida o un error de programación. Ferri le mostró este resultado al personal del INEC para que se hiciera una investigación ulterior.

Otra prueba importante es comparar las frecuencias de imputación con las frecuencias de los datos no corregidos. Los valores asignados (imputados) a una variable deben tener una distribución que se asemeje a la distribución de la variable misma. Si esto no fuera verdad, el programa de crítica e imputación está distorsionando los datos e introduciendo un sesgo. Esto puede ser causado por una especificación incorrecta en la imputación dinámica o por un error de programación. Los resultados se muestran en las Tablas A09 a A22 en el Adjunto A. Estos resultados muestran que las frecuencias de imputación están ciertamente cerca de las frecuencias naturales y, por lo tanto, las imputaciones no están introduciendo ningún sesgo.

5.4. Resumen de Observaciones y Recomendaciones

- El INEC está utilizando correctamente la metodología de crítica e imputación.
- Los especialistas en la materia y los programadores trabajan en conjunción; hay buenas especificaciones de crítica y los programas están bien escritos.
- Los programadores del INEC tienen un muy buen conocimiento de las técnicas de programación de crítica e imputación, así como de la programación de aplicaciones de lote de CSPro.
- Las especificaciones estructurales son sólidas y la dificultad de lidiar con datos a nivel de vivienda y a nivel de hogar ha sido satisfactoriamente resuelta.
- El programa de crítica e imputación todavía no está terminado, como lo demuestra los errores obtenidos en la segunda corrida de los archivos de Managua y Nueva Segovia. El INEC debe resolver estos problemas.
- El INEC debe rutinariamente correr el programa de crítica e imputación dos veces para cada archivo de datos.
- El INEC debe crear un diccionario de datos definitivo para probar el programa de crítica, para definir los valores válidos de cada variable después de la imputación, prestando atención especial a aquellas variables que puedan o no aceptar valores en blanco. Se debe correr el módulo de frecuencias de CSPro (*Frequencies*) incluyendo los valores no válidos.
- El INEC debe correr frecuencias para todas las variables antes y después de correr el programa de crítica e imputación para verificar doblemente que las imputaciones no hayan sesgado los datos.

6. Tabulación

6.1. Plan de Tabulación

Las tablas del censo están divididas en varios volúmenes:

- Unidades de Vivienda Volumen I – 9 tablas por Departamento.
- Unidades de Vivienda Volumen II – 7 de las mismas tablas por Municipio.
- Hogares Volumen I – 13 tablas, algunas por Departamento y otras por Municipio.
- Población Volumen I – 17 tablas sobre características generales por Departamento.
- Población Volumen II – 9 tablas sobre educación por Departamento.
- Población Volumen III – 12 tablas sobre economía por Departamento
- Población Volumen IV – 12 tablas, algunas seleccionadas de otros 3 volúmenes, por Municipio.

Ferri examinó las especificaciones del plan de tabulación preparadas por los expertos en la materia. Están organizadas en hojas electrónicas EXCEL, con una hoja por cada volumen mencionado anteriormente. Cada especificación de tabla representa una hoja dentro de la hoja electrónica. Las especificaciones son bien completas y están bien hechas. Las mismas definen meticulosamente el diseño de la tabla incluyendo los títulos de las columnas así como los talones de texto, el tamaño de la página, los márgenes y la orientación, y también el tipo de letra que se debe usar. Todas las variables involucradas se especifican claramente, así como las reglas de tabulación y los universos correspondientes.

Como en el caso de la crítica e imputación, es muy importante que los especialistas en la materia y los programadores colaboren mutuamente. Hay un proceso iterativo entre el desarrollo del sistema de tabulación y el sistema de crítica e imputación.

Se corren los programas de tabulación con datos reales y se envían a los especialistas en la materia para que revisen las tablas producidas. Hay a veces pequeñas discrepancias y anomalías que deben investigarse. Esto generalmente conduce a un ajuste en las especificaciones de crítica e imputación y el proceso se repite. Ferri fue testigo de la colaboración estrecha entre los especialistas en la materia y los programadores durante su visita. También observó que el INEC está llevando a cabo el proceso iterativo recientemente descrito.

6.2. Programas de Tabulación

Ferri examinó algunos de los programas de tabulación desarrollados por el INEC. Fueron desarrollados usando el módulo CENTS del paquete IMPS. El INEC eligió usar IMPS porque el componente de tabulación de CSPro acababa de hacerse público y no hubo tiempo para aprenderlo y usarlo en trabajos de producción de tablas. El INEC tiene planeado usar CSPro para futuras tabulaciones.

Los programas escritos en CENTS están bien escritos y eso demuestra que se tiene conocimiento del sistema. Los programas usan correctamente el sistema de procesamiento de CENTS para crear tablas a nivel de Departamento y de Municipio.

Ferri verificó la programación de las tablas en CENTS creando algunas de las mismas tablas en CSPro. Ferri corrió el programa de tabulación usando los datos de Managua. En general, ambos sistemas produjeron los mismos números. En algunos casos hubo discrepancias, pero éstas surgieron porque el archivo de datos no estaba completamente limpio. La cláusula para la categoría “Otros” de la orden RECODE (RECODIFICAR) de CENTS estaba recolectando valores no válidos. Por ejemplo, esto causó que algunos registros con blancos en la variable sexo fueran tabulados como mujeres en CENTS, mientras que en CSPro estos casos estarían excluidos de la tabulación. Esto no será un problema cuando los datos estén limpios, pero señala la necesidad de estar absolutamente seguro de que los datos estén limpios, como se describió en la sección anterior.

Ferri trabajó con uno de los programadores del INEC que está experimentando con el módulo de tabulación de CSPro. Ferri recomendó algunas técnicas de tabulación, le explicó ciertas limitaciones y le mostró algunas de las características que todavía no han sido documentadas. Ferri quedó muy impresionado de la facilidad con que el programador entendió el módulo teniendo poca experiencia en el uso del mismo. El programador le mostró a Ferri un defecto en el sistema y Ferri le entregó una actualización que no incluía el defecto con la ayuda de los programadores del Buró de Censos en Washington.

6.3. Resumen de Observaciones y Recomendaciones

- Los especialistas en la materia y los programadores cooperan mutuamente muy bien; las especificaciones del plan de tabulación son buenas y los programas están bien escritos.
- El personal de programación del INEC tiene un muy buen conocimiento de las técnicas de programación de planes de tabulación usando el módulo CENTS de IMPS.
- El personal de programación del INEC debe continuar aprendiendo el uso de CSPro para llevar a cabo planes de tabulación. Deben correr los programas de tabulación en CENTS y comparar los resultados con los obtenidos en CSPro.

7. Productos para la Difusión de Datos

Luego de haber completamente limpiado los datos y de haber preparado el plan de tabulación, hay que desarrollar los productos para la difusión de los datos censales. El INEC todavía no ha llegado a esta etapa. Ferri observó que existe dentro del INEC la capacidad de producir buenos productos para la difusión de la información censal. El INEC ya tiene una página en la red que indudablemente será usada para difundir datos censales. El INEC debería considerar los siguientes productos para la difusión de datos censales:

- Resumen de datos en la internet. El INEC puede publicar las tabulaciones del censo en su sitio en la red. Puede también proporcionar indicadores selectos de interés general para los usuarios de datos. Deberían considerar el uso de hojas factuales por departamento, o municipio, en base a resultados censales de gran importancia.
- Archivos de Uso Público. El INEC puede producir archivos con microdatos con los cuales los usuarios pueden producir sus propias tabulaciones y llevar a cabo sus propios análisis. Hay paquetes de dominio público que el INEC puede usar para producir archivos de uso público, como CSPro y REDATAM+SP. Este último es distribuido por CELADE (Centro Latinoamericano y Caribeño de Demografía). El INEC tiene experiencia en el uso de ambos paquetes.
- Sistema de Recuperación de Tablas (*Table Retrieval System* o *TRS*). El INEC puede organizar todas las tablas publicadas en una base de datos que le permite a los usuarios recuperarlas por área geográfica (departamento y municipio) y por contenido. CSPro contiene una herramienta con la cual se puede crear una base de datos y otras herramientas que pueden ser distribuidas a los usuarios sin costo alguno. La base de datos puede distribuirse en un CD-ROM o a través de la internet.
- Mapas Temáticos. El INEC puede presentar datos censales en mapas temáticos. Estos mapas simples, pero muy efectivos, sirven para mostrar espacialmente datos censales y pueden ser preparados sin la necesidad de contar con un software de SIG complejo y costoso. Las herramientas de dominio público, como CSPro y REDATAM+SP, tienen la capacidad de proporcionar al usuario de datos la posibilidad de crear sus propios mapas temáticos.