

Distr.
RESTRINGIDA

LC/R.2105 101
02 de octubre de 2003

ORIGINAL: ESPAÑOL

CEPAL

Comisión Económica para América Latina y el Caribe

11^o TALLER REGIONAL

METODOLOGÍAS PARA LA INTEGRACIÓN DE BASES DE DATOS DE ENCUESTAS DE HOGARES

Lima, Perú, 16 al 18 de julio de 2003



900029911 - BIBLIOTECA CEPAL

Las opiniones expresadas en este documento, que no ha sido sometido a revisión editorial, son de la exclusiva responsabilidad de los autores y pueden no coincidir con las de la Organización.

03-10-646

ÍNDICE

	<u>Página</u>
PRESENTACIÓN	3
AGENDA	7
PRINCIPALES RESULTADOS Y LOGROS DEL PROGRAMA MECOVI (<i>BID - Banco Mundial - CEPAL</i>)	9
RESUMEN Y CONCLUSIONES	21
PROCESAMIENTO DE DATOS DE LA ENCUESTA PERMANENTE DE HOGARES. <i>Instituto Nacional de Estadística y Censos (INDEC) - ARGENTINA</i>	31
PARTICIPACIÓN DE LA INFORMÁTICA EN EL DISEÑO, DESARROLLO Y EXPLOTACIÓN DE ENCUESTAS A HOGARES. <i>María Teresa Arriola, Graciela Canessa, Raquel Salva y Elba Tonini, Instituto Nacional de Estadística y Censos (INDEC) – ARGENTINA</i>	41
ENCUESTA DE HOGARES. <i>Freddy Ochoa Cabrera, Instituto Nacional de Estadística (INE) - BOLIVIA</i>	49
SISTEMA DE INFORMACIÓN MULTIDIMENSIONAL. ENCUESTA NACIONAL DE EMPLEO. <i>Instituto Nacional de Estadística (INE) - CHILE</i>	57
INFORME METODOLÓGICO INFORMÁTICO ENCUESTA DE HOGARES, AÑO 2003. <i>Departamento Administrativo Nacional de Estadística de Colombia (DANE) - COLOMBIA</i>	77
ENCUESTAS DE HOGARES DE PROPÓSITOS MÚLTIPLES. <i>Carlos Loría Loría, Instituto Nacional de Estadística y Censos (INEC) – COSTA RICA</i>	85
ASPECTOS METODOLÓGICOS DE LA ENCUESTA DE HOGARES. <i>Oficina Nacional de Estadística (ONE), Centro de Estudios de Población y Desarrollo (CEDPE) - CUBA</i>	101
ENCUESTA DE HOGARES – SITUACIÓN ACTUAL. <i>Carlos Obando, Instituto Nacional de Estadística y Censo (INEC) - ECUADOR</i>	115
METODOLOGÍA PARA LA INTEGRACIÓN DE BASES DE DATOS DE LA ENCUESTA NACIONAL DE EMPLEO E INGRESOS - ENEI. <i>Instituto Nacional de Estadística (INE) - GUATEMALA</i>	123
ENCUESTA NACIONAL DE EMPLEO E INGRESOS – ENEI 2002. <i>Instituto Nacional de Estadística (INE) - GUATEMALA</i>	133
PROCESAMIENTO DE DATOS DE LA ENCUESTA PERMANENTE DE HOGARES DE PROPÓSITOS MÚLTIPLES. <i>Instituto Nacional de Estadística (INE) - HONDURAS</i>	141
DATA WAREHOUSE PARA LA PRESTACIÓN DEL SERVICIO PÚBLICO DE INFORMACIÓN ESTADÍSTICA. <i>Ricardo Luján Salazar, Instituto Nacional de Estadística Geográfica e Informática (INEGI) - MÉXICO</i>	147
ENIGH: VALIDACIÓN Y CONSTRUCCIÓN DE LA BASE DE DATOS. <i>Benito Durán, Instituto Nacional de Estadística Geográfica e Informática (INEGI) - MÉXICO</i>	157
PROCESAMIENTO DE DATOS DE LA ENCUESTA DE MEDICIÓN DEL NIVEL DE VIDA. <i>Instituto Nacional de Estadística y Censo (INEC) - NICARAGUA</i>	163

INFORME METODOLÓGICO INFORMÁTICO ENCUESTA DE HOGARES AÑO 2002. <i>Marcela Peñalba García, Dirección de Estadística y Censo (DEC) - PANAMÁ</i>	175
ENCUESTA INTEGRADA DE HOGARES 2000/1 PROCESAMIENTO DE DATOS. <i>Oscar Esteban Correa, Dirección General de Estadística, Encuestas y Censos (DGEEC) - PARAGUAY</i>	187
METODOLOGÍA PARA LA INTEGRACIÓN DE BASES DE DATOS DE ENCUESTAS DE HOGARES. <i>Instituto Nacional de Estadística e Informática (INEI) - PERÚ</i>	213
DIFUSIÓN DE INFORMACIÓN DE LA ENCUESTA NACIONAL DE HOGARES. <i>Instituto Nacional de Estadística e Informática (INEI) - PERÚ</i>	245
METODOLOGÍA PARA LA INTEGRACIÓN DE BASES DE DATOS DE ENCUESTAS DE HOGARES. <i>Banco Central – REPÚBLICA DOMINICANA</i>	253
METODOLOGÍAS PARA LA INTEGRACIÓN DE BASES DE DATOS DE ENCUESTAS DE HOGARES. <i>Carlos E. Calvo, Instituto Nacional de Estadística (INE) - URUGUAY</i>	261
METODOLOGÍA PARA LA INTEGRACIÓN DE BASES DE DATOS DE ENCUESTAS DE HOGARES. <i>Instituto Nacional de Estadística (INE) - VENEZUELA</i>	273
ESTANDARIZACIÓN DE LAS BASES DE DATOS DE LAS ENCUESTAS DE HOGARES. <i>CEPAL</i>	287
BADEHOG. <i>CEPAL</i>	293
CENSUS AND SURVEY PROCESSING SYSTEM. <i>SERPRO S.A.</i>	301
LISTA DE PARTICIPANTES.	307

PRESENTACIÓN

El interés por conocer el desempeño de los países y sus progresos en materia económica y social, ha generado un renovado interés de parte de gobiernos y organismos internacionales por disponer de información confiable y oportuna que de cuenta de los avances que se observan en los distintos indicadores que evalúan el nivel de vida de las personas.

En ese sentido, el balance social del primer trienio del nuevo siglo deja ver un panorama poco auspicioso en materia de superación de la pobreza, ya que las cifras más recientes generadas por la CEPAL confirman que la evolución de este fenómeno se estancó, y se estima que actualmente 220 millones de latinoamericanos y caribeños enfrentan alguna situación de pobreza, al tiempo que las proyecciones señalan que esta cifra continuará aumentando para ubicarse en 225 millones al finalizar el presente año.

Ante un entorno internacional altamente cambiante en materia económica y social, los compromisos asumidos por los Jefes de Estado y de Gobierno en la “*Declaración del Milenio*” representan un hito, ya que se percibe una clara decisión de los países por realizar los esfuerzos necesarios para mejorar las condiciones de vida de su población. La determinación de lograr que en el 2015 la tasa de pobreza extrema se reduzca a la mitad, representa uno de los múltiples desafíos que los estados miembros de la Naciones Unidas se plantearon como meta, situación que ha sido acogida con gran beneplácito por la comunidad internacional.

Es en este contexto que se reafirma la necesidad de disponer de información que permita hacer el seguimiento periódico de la manera en que evolucionan los distintos indicadores asociados al bienestar de las personas, y en especial de aquellos que dan cuenta de los niveles de pobreza, así como los que permiten conocer la distribución del ingreso entre los distintos sectores de la población.

Para dar respuesta adecuada y oportuna a estas y otras interrogantes que se han instalado en el debate público, se considera relevante que en los países se continúe apoyando la generación de información estadística a partir de las iniciativas existentes relacionadas con la realización de encuestas de hogares, identificando las acciones que se deben efectuar para reforzar los sistemas nacionales de información como una actividad que debe ubicarse con la más alta prioridad en las agendas nacionales de los países de la región.

Los compromisos internacionales asumidos por los países de América Latina y el Caribe obligan a replantear la necesidad de contar con sistemas de información económica y social confiables y oportunos, que permitan monitorear el efecto de las políticas públicas en el bienestar de las familias. En este sentido, las encuestas de hogares se han consolidado como una herramienta necesaria para apoyar este proceso, y en la actualidad los datos que generan son ampliamente utilizados para evaluar la política pública y apoyar a las distintas instancias vinculadas a la toma de decisiones.

En los años recientes, en la mayoría de los países se pusieron en marcha diversas iniciativas orientadas a fortalecer los sistemas de información, lo cual ha permitido a las oficinas nacionales de estadística de la región consolidar sus programas de encuestas de hogares. No obstante, y a pesar de los importantes avances registrados, aún se perciben necesidades que deben ser atendidas con el objetivo de mejorar la calidad de la información que se recolecta, así como ampliar la cobertura y desagregación geográfica de las muestras y aumentar la periodicidad con

la que se recopila la información, entre otros temas en los cuales se considera necesario que los países deben continuar avanzando.

El desarrollo de la tecnología informática ha significado una gran oportunidad para mejorar los sistemas de recolección de información, la entrada y procesamiento de datos, así como incrementar de manera notable las opciones de disseminación oportuna de los mismos.

Es en este contexto que la iniciativa de asistencia técnica y financiera emprendida por el Banco Interamericano de Desarrollo (BID), el Banco Mundial y la Comisión Económica para América Latina y el Caribe (CEPAL) cobra mayor relevancia, ya que distintas acciones emprendidas por el Programa para el Mejoramiento de las Encuestas y la Medición de las Condiciones de Vida en América Latina y el Caribe (MECOVI), han permitido consolidar las experiencias nacionales, abriendo la posibilidad de identificar buenas prácticas y señalar las áreas en las que se pueden mejorar los diseños estadísticos, a la vez que ampliar las opciones para hacer un mejor uso de la información generada por las encuestas en hogares.

Desde su concepción inicial, el programa MECOVI consideró la participación de la CEPAL en distintas actividades asociadas al desarrollo de esta iniciativa, especialmente en aquellas acciones orientadas a diseñar y coordinar la ejecución de un conjunto de talleres regionales, concebidos como un espacio de reflexión y análisis en el cual los países intercambian experiencias y profundizan el debate sobre temas relevantes y de interés común, vinculados a las actividades sustantivas del programa y a las experiencias nacionales que se constituyen en buenas prácticas que merecen ser analizadas, compartidas y ampliamente difundidas.

La historia de los talleres del MECOVI se remonta a la segunda mitad de la década de los noventa. Entre 1998 y 2001 se realizaron 8 talleres regionales en los cuales se abordaron los siguientes temas: Planificación y Desarrollo de Encuestas de Hogares para la Medición de las Condiciones de Vida; Medición del Ingreso en las Encuestas de Hogares; Medición del Gasto en las Encuestas de Hogares; La medición de la Pobreza: El Método de las Líneas de Pobreza; e Indicadores sobre el Desarrollo Social. En todos los casos, a la reflexión teórica y conceptual se le incorporaron experiencias nacionales que aportaron evidencia empírica de gran valor que han permitido mejorar las prácticas nacionales, identificando las bondades y limitaciones de las distintas opciones metodológicas que fueron analizadas y discutidas en cada oportunidad, e identificando áreas para continuar mejorando los aspectos técnicos del diseño conceptual, así como los mecanismos e instrumentos de recolección de información. Debido a la importancia que el programa MECOVI le confiere al perfeccionamiento de las fuentes de información, el séptimo Taller Regional se centró en discutir los aspectos conceptuales y técnicos asociados al Diseño e Implementación de un Sistema Integrado de Encuestas de Hogares (SIEH).

Por su parte, la agenda de trabajo del Octavo taller surgió del interés manifestado por los países beneficiarios y las instituciones co-patrocinadoras del programa, ante la necesidad de hacer un alto en el camino para reflexionar y evaluar los beneficios que ha generado el MECOVI, a partir del análisis de las distintas experiencias de los países que se han beneficiado directamente de los recursos financieros, evaluando su contribución al fortalecimiento institucional, los avances acerca de los determinantes de la pobreza, así como de las acciones que se están efectuando para que los Sistemas Integrados de Encuestas de Hogares se consoliden como un patrimonio de los Institutos Nacionales de Estadística (INE's).

Las agendas de trabajo diseñadas para el noveno y décimo taller se elaboraron con la clara intención de generar sinergia entre los temas que se abordaron, debido a su vinculación con la

ronda de los Censos de Población y Vivienda del 2000. Se consideró oportuno aprovechar las recientes experiencias desarrolladas en la actualización y construcción de los marcos de muestreo, así como en la renovación de los diseños estadísticos de las muestras que dan sustento a los programas permanentes de encuestas en hogares.

Finalmente, los temas que se analizaron en la undécima edición de los talleres del MECOVI permitieron incorporar a la agenda de discusión una temática de gran relevancia. Se considera imprescindible que la tecnología informática se ponga al servicio de los institutos de estadísticas, y de manera especial en lo que se refiere al aprovechamiento de los avances computacionales para mejorar el diseño y explotación de las bases de datos que se forman a partir de la información que recopilan las encuestas de hogares.

En ese sentido, se decidió que el tema central del taller se orientara a analizar *las “Metodologías para la integración de bases de datos de encuestas de hogares”*, ya que este tema se considera como muy relevante, ante la creciente utilización que se está haciendo de los resultados generados por las encuestas para monitorear la eficacia de las políticas públicas y medir sus efectos en el bienestar individual y colectivo.

Al igual que en ocasiones anteriores, los materiales contenidos en este informe se encuentran a disposición de los interesados en el sitio de la CEPAL: www.eclac.cl.

**PROGRAMA PARA EL MEJORAMIENTO DE LAS ENCUESTAS Y LA MEDICION DE LAS
CONDICIONES DE VIDA EN AMERICA LATINA Y EL CARIBE (MECOVI)
BID - BANCO MUNDIAL - CEPAL**

**11° TALLER REGIONAL
METODOLOGÍAS PARA LA INTEGRACIÓN DE BASES
DE DATOS DE ENCUESTAS DE HOGARES
Lima, Perú, 16 al 18 de julio de 2003**

INEI

CEPAL

Agenda

Miércoles 16	
09:00 – 09:30	Registro de participantes
09:30 – 10:00	Palabras de bienvenida (<i>INEI, CEPAL</i>)
10:00 – 10:15	Presentación de la Agenda y de los objetivos del Taller (<i>CEPAL</i>)
TEMA 1	Experiencias Nacionales
10:15 – 11:00	Argentina
11:00 – 11:30	<i>Café</i>
11:30 – 12:00	Bolivia
12:00 – 12:30	Chile
12:30 – 13:00	Discusión
13:00 – 15:00	<i>Almuerzo</i>
15:00 – 15:30	Colombia
15:30 – 16:00	Costa Rica
16:00 – 16:30	Ecuador
16:30 – 16:45	<i>Café</i>
16:45 – 17:15	Guatemala
17:15 – 18:00	Discusión
Jueves 17	
TEMA 1 (...)	(continuación)
09:00 – 09:30	Honduras
09:30 – 10:00	México
10:00 – 10:30	Nicaragua
10:30 – 11:00	Panamá
11:00 – 11:30	<i>Café</i>

11:30 – 12:00	Paraguay
12:00 – 13:00	Discusión
13:00 – 15:00	<i>Almuerzo</i>
15:00 – 15:45	Perú
15:45 – 16:15	República Dominicana
16:15 – 16:30	<i>Café</i>
16:30 – 17:00	Uruguay
17:00 – 17:30	Venezuela
17:30 – 18:00	Discusión
Viernes 18	
TEMA 2	Visión de los proveedores de paquetes estadísticos
09:00 – 10:00	CsPro
10:00 – 11:00	SPSS
11:00 – 11:30	<i>Café</i>
TEMA 3	Experiencia de la CEPAL
11:30 – 12:30	La experiencia de la CEPAL en la sistematización de bancos de datos de las encuestas de hogares
12:30 – 13:00	Síntesis y Conclusiones
13:00 – 13:15	Clausura

**PRINCIPALES RESULTADOS Y LOGROS DEL
PROGRAMA MECOVI
(diciembre 1996-julio 2003)**

BID-BANCO MUNDIAL-CEPAL

ÍNDICE

	<u>Página</u>
A. Actividades en Países Participantes.....	11
B. Actividades Regionales.....	17
C. Incorporación de Nuevos Países	18
D. Evaluación	19
E. Coordinación.....	19
F. Conferencia Estadística de las Américas (CEA).....	19

- 1.1 A continuación se presenta un breve resumen de los reportes de avances de actividades por país preparados para el Comité Directivo (Steering Committee) y el Consejo Asesor (Advisory Board) del Programa MECOVI.
 - 1.2 El Programa MECOVI tiene dos componentes principales: (a) actividades en el ámbito de países participantes; y (b) actividades a escala regional (países de América Latina y el Caribe).
- A. Actividades en Países Participantes**
- 1.3 A la fecha los países participantes, a través de las oficinas nacionales de estadística, son Argentina, Bolivia, El Salvador, Guatemala, Nicaragua, Paraguay y Perú. Se están finalizando las actividades de coordinación para consolidar la incorporación al Programa de Ecuador, Honduras y República Dominicana. La asistencia técnica otorgada por el Programa a los institutos nacionales de estadística de estos países ha mejorado las encuestas de hogares en términos de su cobertura temática y geográfica. En cuanto al contenido de las encuestas de hogares, se ha expandido la cobertura temática de los cuestionarios tradicionales para incorporar módulos nuevos en algunos o todos los países beneficiarios, incluyendo la medición de la nutrición de los niños (antropometría), remesas familiares, actividad agropecuaria, uso del tiempo de los miembros del hogar, capital social, acceso a programas sociales, seguridad ciudadana, y manejo de los choques adversos de los hogares (ver Cuadro 1). Se ha ampliado la cobertura geográfica de las encuestas al área rural por primera vez en algunos países (Paraguay, Bolivia), en donde justamente se presenta una mayor incidencia y severidad de la pobreza.
 - 1.4 *Asistencia técnica.* El Programa MECOVI ha otorgado asistencia técnica, por medio de consultores nacionales e internacionales expertos en encuestas de hogares, sobre los siguientes aspectos: i) actualización de la cartografía y el diseño de la muestra; iii) desarrollo de cartografía digitalizada; iv) organización del trabajo de campo; v) organización de la entrada de datos de modo descentralizado; vi) desarrollo de los principales criterios de consistencia y control de la calidad en el campo; vii) organización de la base de datos; viii) procesamiento y análisis de la información.
 - 1.5 *Interacción entre usuarios y productores de información.* Se han promovido mecanismos para lograr un mayor estrechamiento de la relación entre productores y usuarios de encuestas de hogares a fin de asegurar que las encuestas capten la información requerida por el país. Para tal efecto, en todos los países participantes se ha conformado un Comité Interinstitucional, compuesto por representantes de diversos organismos públicos y privados, productores y usuarios de encuestas de hogares.
 - 1.6 *Capacitación.* Las oficinas de estadística de los países participantes han organizado talleres de trabajo y cursos de capacitación, con la participación de productores y usuarios de encuestas, para discutir tanto aspectos de metodología y como de resultados de los estudios realizados con la información de las encuestas de hogares. Los temas de los cursos han sido muestreo, uso de software para procesar encuestas, aspectos conceptuales sobre condiciones de vida, etc.
 - 1.7 *Actividades por país.* A continuación se presentan los cuadros resumen por país ordenados cronológicamente en base al inicio de las actividades del Programa.

1.8 Paraguay.

Organismo Ejecutor	Dirección General de Estadísticas, Encuestas y Censos (DGEEC)
Inicio de actividades	Las actividades de MECOVI se iniciaron a fines de 1996.
Operación de campo	Encuesta Integrada de Hogares (EIH) a nivel nacional (urbana y rural) de agosto de 1997 a agosto de 1998. Encuesta Permanente de Hogares (EPH), nacional, de agosto a diciembre 1999. Actualmente la EIH 2000-01 está siendo procesada.
Cuestionarios	Cuestionarios de hogares, a centros poblados rurales, de precios, y sobre antropometría.
Base de datos	Las bases de datos de la EIH 1997-98, EPH 1999 y EIH 2000-01 están disponibles para los usuarios.
Asistencia técnica	La asistencia técnica ha contribuido a mejorar: i) el contenido temático de la encuesta; ii) la actualización de la cartografía y el diseño de la muestra; iii) el desarrollo de la cartografía digitalizada; iv) la organización del trabajo de campo; v) la entrada de datos de modo descentralizado; vi) el desarrollo de los principales criterios de consistencia y control de la calidad en el campo; vii) la organización de la base de datos; viii) el procesamiento y análisis de la información.
Capacitación y talleres	Se han organizado cursos de entrenamiento en el manejo de software para el procesamiento de datos, muestreo y análisis de encuestas. Además, los técnicos se han visto beneficiados mediante la capacitación en el terreno con la presencia de los consultores nacionales e internacionales, y los talleres de trabajo realizados.
Estudios y publicaciones	Se han realizado publicaciones y estudios utilizando la información de las encuestas de 1995, 1996 y 1997/98: i) edición de documentos de Divulgación Popular (No. 1 y 2); ii) elaboración del estudio sobre el perfil de condiciones de vida en 1995; iii) perfil de condiciones de vida en 1996; iv) estudios descriptivos sobre la situación de la fuerza laboral en 1996, perfil de la salud 1996, y perfil educativo 1996; v) perfil de condiciones de vida 1997/98; vi) revista "Economía & Sociedad" números 1 al 5 y, vii) folletos informativos.
Comité Interinstitucional	El Comité está formado por representantes de: Secretaría Técnica de Planificación de la Presidencia de la República, Banco Central de Paraguay, Centro de Documentación y Estudio (CDE), Universidad Nacional de Asunción, y la DGEEC.
Fondo de Estudios	El Fondo de Estudios se puso en marcha en 1999, utilizando la información de la EIH 1997-98. Se han financiado tres estudios en 1999. La segunda ronda del Fondo de Estudios ha concluido. Los estudios preparados han sido publicados como parte de la revista "Economía & Sociedad".
Evaluación	Durante 1998 se realizó una evaluación de los cuestionarios ejecutados (hogares y comunitario). En el IV trimestre de 2000 se evaluó, con excelentes resultados, el proceso de implementación de la asistencia técnica otorgada.

1.9 Perú.

Organismo Ejecutor	Instituto Nacional de Estadística e Informática (INEI)
Inicio de actividades	En Perú se iniciaron las operaciones del Programa MECOVI en Febrero de 1997.
Operación de campo	Se han completado las rondas del IV trimestre de 1997, del II, III y IV trimestres de 1998, del I, II, III y IV trimestres de 1999, y del I, II, III y IV trimestres de 2000 de la "Encuesta Nacional de Hogares" (ENAHO) con cobertura nacional (urbana y rural). Las encuestas correspondientes a 2002 están siendo procesadas.
Cuestionarios	Cuestionarios de hogares, a comunidades rurales, y de precios rurales.
Base de datos	La base de datos de las ENAHO del IV trim. 1997, los trimestres I, II, III y IV de 1998, 1999 y 2000 están disponibles para los usuarios. Además de las bases del III y IV trimestres del 2001
Asistencia técnica	La asistencia técnica ha estado orientada a: I) mejorar el contenido temático del cuestionario mediante el análisis de las principales variables usando la ENAHO 1996 (gasto, ingreso, empleo, acceso a servicios sociales básicos); ii) el diseño de la muestra; iii) mejorar los métodos de capacitación de los encuestadores y supervisores (de modo descentralizado); iv) el desarrollo de sistemas de control de calidad en todas las etapas de la encuesta, incluyendo un banco de preguntas y respuestas para compartir los problemas y las soluciones sobre aspectos de la encuesta en el Intranet del INEI; v) mejoramiento del análisis y procesamiento de los datos.
Capacitación y talleres	Se ha reforzado las capacidades de los técnicos del INEI y los usuarios de las encuestas mediante cursos nacionales de capacitación en las siguientes áreas: manejo de software, uso de la base de datos y aplicaciones de la ENAHO, muestreo, análisis de datos. Estos se han complementado mediante el desarrollo de talleres y seminarios metodológicos, a través de la capacitación en el terreno con la presencia de los consultores nacionales e internacionales y mediante cursos y talleres regionales. El INEI ha organizado dos de dichos talleres: el primero, para discutir metodologías utilizadas en común en los países participantes en el MECOVI; el segundo, para examinar el uso de información de encuestas de hogares de condiciones de vida en conjunto con información de censos.
Estudios y publicaciones	Se han elaborado estudios sobre el perfil de la pobreza y distribución del ingreso en el Perú con base a la información de la ENAHO 1996, 1997 y 1998. Asimismo, se han realizado un estudio para definir las líneas de pobreza en Perú y estudios sobre el perfil de condiciones de vida en 1997 y 1998. Además se han hecho publicaciones descriptivas de los sectores sociales (educación, vivienda, salud, empleo, etc.). Se publicaron los documentos "La pobreza en el Perú 2001", "Condiciones de Vida en el Perú: Evolución 1997-2001" y "Condiciones de Vida en los Departamentos del Perú, 2001".
Comité Interinstitucional	El Comité está formado por representantes de: Ministerios de la Presidencia, de Trabajo, de Salud, de Educación, de Transportes, Vivienda y Construcción, Promoción de la Mujer y Desarrollo Humano, Economía; Banco Central de Reserva; Fondo de Compensación para el Desarrollo Social (FONCODES); Instituto Nacional de Salud; Universidad del Pacífico; Grupo de Análisis para el Desarrollo (GRADE); Asociación Laboral para el Desarrollo (ADEC-ATC) e INEI.
Fondo de Estudios	El uso del Fondo de Estudios fue activado en 1999, financiando 16 proyectos de investigación. En la segunda convocatoria se han seleccionado 16 nuevos estudios.
Evaluación	Durante 1998 se realizó una evaluación de los cuestionarios ejecutados (hogares y comunitario), en el IV trimestre de 2000 se evaluó, con excelentes resultados, el proceso de implementación de la asistencia técnica otorgada.

1.10 *El Salvador.*

Organismo Ejecutor	Dirección General de Estadística y Censos (DIGESTYC).
Inicio de actividades	Septiembre de 1997.
Operación de campo	La operación de campo de la Encuesta de Hogares de Propósitos Múltiples (EHPM), a nivel nacional (urbano y rural), se ejecutaron entre mayo - diciembre de 1998 y febrero - diciembre 1999. La encuesta de 2000 está siendo procesada, en tanto que la ejecución de la de 2001 se vio retrasada como consecuencia de los terremotos de enero y febrero de 2001. La encuesta de 2002 está en campo.
Cuestionarios	Cuestionarios de hogares, comunitario, precios rurales, y antropometría.
Base de datos	Las bases de datos de la EHPM 1998, 1999, 2000 y 2001 están disponibles.
Asistencia técnica	La asistencia técnica ha estado dirigida a: i) actualización del marco muestral durante el periodo septiembre 1997- marzo 1998; ii) mejorar el contenido temático de los cuestionarios; iii) diseño de la muestra; iv) organizar el trabajo de campo y la entrada de datos; v) sistematizar métodos de control de calidad; vi) definir un plan de trabajo orientado a la digitalización de la cartografía; vii) uso de software para el manejo de bases de datos (SPSS, ISSA); y, viii) recolección y manejo de información antropométrica.
Capacitación y talleres	Se han realizado talleres de trabajo de tipo metodológico para discutir la definición de los cuestionarios, muestra y control de calidad, con la participación de consultores nacionales e internacionales. Se han realizado cursos en el uso de software para manejo de datos, sobre antropometría, y análisis de encuestas de hogares.
Estudios y publicaciones	Estimados mensuales de la tasa de desempleo. Estudios descriptivos sobre nutrición, condiciones de vida y tabulados generales. Se preparó un perfil de pobreza de la década de los noventas.
Comité Interinstitucional	El Comité está conformado por representantes de: Ministerios de Salud, de Educación, de Relaciones Exteriores, y de Trabajo; Banco Central de Reserva; FUSADES; Fondo de Inversión Social (FISDL); Universidad José Simeón Cañas; DIGESTYC.
Fondo de Estudios	El uso del Fondo de Estudios fue activado en 1999 con la selección de cuatro estudios. Actualmente se está convocando a una segunda selección de estudios.
Evaluación	Durante 1998 se realizó una evaluación de los cuestionarios ejecutados (hogares y comunitario). En el IV trimestre de 2000 se evaluó, con excelentes resultados, el proceso de implementación del programa.

1.11 *Argentina.*

Organismo Ejecutor	Instituto Nacional de Estadística y Censos (INDEC).
Inicio de actividades	Noviembre 1998.
Operación de campo	Ejecución de la "Encuesta Permanente de Hogares" (EPH), dos veces por año. Una versión revisada de esta encuesta será ejecutada de modo continuo. Se está procesando el Encuesta de Condiciones de Vida 2001.
Cuestionarios	Cuestionario de hogares básico (ingreso y empleo). Cuestionario ampliado sobre condiciones de vida (a ser implementado en 2001). Cuestionario reformulado de la EPH.
Base de datos	Disponibles tres meses después de terminar el trabajo de campo.

Asistencia técnica	En proceso: (i) actualización del marco muestral; (ii) rediseño del cuestionario de la EPH; (iii) estrategia para un sistema integrado de encuestas de hogares; (iv) módulo de seguimiento del consumo; (v) revisión de la medición del ingreso; (vi) indicadores sociodemográficos; (vii) diseño de publicaciones; (viii) indicadores laborales; (ix) seguimiento del consumo; y (x) armonización de los conceptos y definiciones de la encuesta y los registros administrativos.
Capacitación y talleres	Se han realizado dos cursos nacionales sobre muestreo, un curso sobre indicadores sociales, y dos talleres nacionales. En marzo del 2000 se llevó a cabo el tercer curso regional de capacitación en el INDEC con la participación de 35 estudiantes de 18 países. En abril de 2001 se realizó el quinto curso regional del MECOVI en el INDEC, se contó con la asistencia de 30 técnicos de 18 países. En abril de 2002 se llevó a cabo el séptimo curso regional del MECOVI en el INDEC, en el cual participaron 30 técnicos de 18 países. El noveno curso regional del MECOVI está se llevará a cabo en el INDEC en mayo de 2003.
Estudios y publicaciones	(i) Estudio de los gastos e ingresos de los hogares (ENGHO 1996-97); (ii) análisis del mercado laboral; y, (iii) educación.
Comité Interinstitucional	En proceso de formación.
Fondo de Estudios	El fondo fue activado en 1999. Actualmente se están financiando 11 propuestas de investigación (6 de investigadores expertos y 5 de estudiantes de posgrado). En la convocatoria de 2000 se financiaron ocho proyectos (cuatro de investigadores expertos y cuatro de estudiantes de posgrado).
Evaluación	En mayo de 2002 se realizó una evaluación de medio término del programa.

1.12 Nicaragua.

Organismo Ejecutor	Instituto Nacional de Estadística y Censos (INEC).
Inicio de actividades	Mayo 1998
Operación de campo	"Encuesta de Medición de Niveles de Vida" (EMNV), en áreas urbanas y rurales: mayo – agosto 1998. Encuesta de seguimiento del impacto del Huracán Mitch: 1999. Se finalizó el trabajo de campo de la EMNV-2001, se está procesando la base de datos.
Cuestionarios	Cuestionarios de hogares, comunitario, precios rurales, y antropometría.
Base de datos	La base de datos de la EMNV 1998, 1999 y 2001 está disponible para los usuarios
Asistencia técnica	(i) Diseño de cuestionario; (ii) diseño de muestra; (iii) organización del trabajo de campo; (iv) organización de la base de datos; (v) procesamiento y análisis de la información; (vi) generación del agregado de consumo; y (vii) construcción de un mapa de pobreza.
Capacitación y talleres	Entrenamiento en el uso de SPSS. En 1999 se realizaron talleres sobre muestreo y metodología de la encuesta de hogares.
Estudios y publicaciones	Folleto de divulgación popular y las principales tabulaciones. Informe general de resultados de la EMNV 1998-99. Estudio sobre pobreza rural, y un mapa de pobreza para 2001.
Comité Interinstitucional	Miembros: Instituto Nacional de Estadística y Censos (INEC); Ministerio de Economía y Desarrollo (MEDE); Ministerio de Educación (MED); Ministerio de Salud (MINSAL); Ministerio de Trabajo (MITRAB); Instituto de fomento Municipal (INIFOM); Ministerio de Agricultura y Ganadería (MAG) ; Banco Central de Nicaragua (BCN); Fondo de Inversión Social de Emergencia (FISE); Ministerio de Acción Social (MAS); Programa de las Naciones Unidas para el Desarrollo (PNUD); Fondo de Población de las Naciones Unidas (FNUAP); y, Banco Mundial.
Fondo de Estudios	
Evaluación	Se realizó entre octubre y noviembre de 2000, con buenos resultados.

1.13 *Bolivia.*

Organismo Ejecutor	Instituto Nacional de Estadística (INE).
Inicio de actividades	Mayo 1999
Operación de campo	Encuesta Continua de Hogares (ECH), noviembre 1999, y Encuesta Continua de Hogares IV trimestre 2000 y IV trimestre 2001.
Cuestionarios	Cuestionarios de hogares.
Base de datos	Disponible ECH, 1999, 2000 y 2001.
Asistencia técnica	i) Diseño de cuestionario; ii) diseño de muestra; y, iii) organización de trabajo de campo, y (iv) metodología para la consistencia de datos.
Capacitación y talleres	Se han programado cursos y talleres para técnicos del INDEC y usuarios externos.
Estudios y publicaciones	Informe de resultados de la ECH 1999.
Comité Interinstitucional	A ser conformado.
Fondo de Estudios	
Evaluación	En marzo de 2002 se llevó a cabo una evaluación de medio término del programa.

1.14 *Guatemala.*

Organismo Ejecutor	Instituto Nacional de Estadística (INE).
Inicio de actividades	Octubre de 1999
Operación de campo	"Encuesta Nacional de Condiciones de Vida" (ENCOVI) (julio – noviembre de 2000) terminada. Se han realizado cuatro rondas de la Encuesta Nacional de Empleo e Ingresos (ENEI).
Cuestionarios	Cuestionarios de hogares, comunitario, precios, y antropometría.
Base de datos	ENCOVI 2000 disponible.
Asistencia técnica	Asistencia técnica en el diseño del cuestionario, actualización del marco muestral, y organización de trabajo de campo. Además, en el análisis de la información y procesamiento y manejo de datos de la "Encuesta Nacional de Ingresos y Gastos Familiares" (ENIGFAM, 1998).
Capacitación y talleres	Por ejecutar.
Estudios y publicaciones	Estudio en proceso sobre condiciones de vida utilizando datos de ingreso y gasto de la "Encuesta Nacional de Ingresos y Gastos Familiares" (ENIGFAM) 1997-98. "Perfil de la Pobreza en Guatemala" basado en la información de la ENCOVI.
Comité Interinstitucional	Ministerio de Salud, Ministerio de Trabajo, Banco de Guatemala, FONAPAZ, INCAP/OPS, SEGEPLAN, GTZ, Fundación Soros, CIEN, Universidad del Valle, Universidad Rafael Landívar, UNFPA, PNUD, UNICEF, USAID, FLACSO, DIACO, Instituto Nacional de Estadística.
Fondo de Estudios	A ser activado en 2003.
Evaluación	Se realizará a los dos años de iniciada la ejecución.

B. Actividades Regionales

- 1.15 Las actividades regionales del Programa MECOVI incluyen: talleres metodológicos, cursos de capacitación y un banco de datos. A la fecha se han realizado las siguientes actividades:
- 1.16 *Talleres regionales.* El Programa MECOVI con la meta de asegurar la continuidad de las actividades de producción de encuestas de hogares de calidad por parte de las oficinas de estadística, ha promovido actividades para crear una masa crítica de técnicos en encuestas de hogares, productores y usuarios, mediante la organización de talleres de trabajo y cursos de capacitación tanto a escala nacional como regional, y a través de la asistencia técnica en el terreno por medio de los expertos en encuestas. A la fecha se han organizado diez talleres de trabajo a nivel regional para discutir aspectos metodológicos de las encuestas de hogares. El primero, "Planificación y desarrollo de encuestas de hogares para la medición de condiciones de vida" (1-3 de abril de 1998 en México). El segundo, "Medición de los ingresos del hogar" (10-13 de noviembre de 1998 en Argentina). El tercero, "Medición del gasto de los hogares" (24-28 de mayo de 1999 en México). El cuarto, "La medición de la pobreza: El método de las líneas de pobreza" (16-19 de noviembre de 1999 en Argentina). El quinto taller, "La medición de la pobreza: Métodos y Aplicaciones (continuación)" (6-8 de junio, 2000, en México). El sexto taller, "Indicadores sobre el desarrollo social" (15-16 de noviembre, 2000, en Argentina). El séptimo taller, "Diseño e Implementación de un Sistema Integrado de Encuestas de Hogares" (30 de mayo al 1ero de junio, 2001, en México). El octavo taller se llevó a cabo en Buenos Aires del 28 al 30 de noviembre, 2001, el tema fue "Pasado, Presente y Futuro del Programa MECOVI". El noveno taller se realizó en la ciudad de Lima, Perú, en coordinación con el Instituto Nacional de Estadística e Informática (INEI) del 17 al 19 de junio, 2002, el tema fue "Diseño y Construcción de Marcos de Muestreo para la Encuestas de Hogares". El décimo taller se realizó en la ciudad de Buenos Aires, Argentina, en coordinación con el Instituto Nacional de Estadística y Censos (INDEC) del 27 al 29 de noviembre, 2003, el tema fue "La Práctica del Muestreo para el Diseño de las Encuestas de Hogares". En todos los talleres participaron representantes, productores y usuarios de la mayoría de países de la región.
- 1.17 *Cursos de capacitación.* Se han realizado ocho cursos de capacitación regional sobre "Diseño, implementación y análisis de encuestas de hogares", de tres semanas de duración cada uno. El primero del 1-17 de marzo de 1999, y el segundo del 30 de agosto al 17 de septiembre de 1999, en conjunto con el Instituto Nacional de Estadística, Geografía e Informática (INEGI) de México. El tercer curso se realizó en cooperación con el Instituto Nacional de Estadística y Censos (INDEC) de Argentina, en Buenos Aires, del 29 de marzo al 18 de abril de 2000. El cuarto tuvo lugar en el INEGI (4 al 22 de septiembre de 2000). El quinto curso se llevó a cabo en cooperación con el INDEC, del 16 de abril al 4 de mayo de 2001 en Argentina. El sexto curso tuvo lugar en el INEGI (3 al 21 de septiembre, 2001). El séptimo curso se organizó en coordinación con el INDEC, del 8 al 26 de abril, 2002. En los siete cursos se contó con la participación de un total de 230 productores y usuarios de encuestas de la mayoría de los países de la región. El octavo curso se llevó a cabo en cooperación con el INEGI (26 de agosto al 13 de septiembre de 2002).
- 1.18 *Fortalecimiento de la Capacidad de Análisis de la Información.* El Programa MECOVI con la finalidad de potenciar un amplio uso y análisis de la información de las encuestas de hogares generadas en los países, ha implementado un Fondo de Estudios con el objetivo de promover y apoyar la realización de investigaciones por profesionales nacionales. En este contexto, a la fecha en Paraguay se han concluido tres estudios correspondientes a la primera ronda, y nueve de la segunda ronda, así como cuatro estudios en El Salvador, 16 estudios en Perú y 19 en Argentina. Los otros países están en proceso de lanzar la convocatoria.

- 1.19 *Mejoramiento del Acceso y Difusión de la Información.* Con ayuda de la asistencia técnica los países han mejorado la organización y documentación de las encuestas de hogares. Además un elemento esencial que se ha promovido es el amplio acceso a la información obtenida, tanto a las bases de datos originales como a los resultados estimados. Las oficinas de estadística están proveyendo esta información a los usuarios del gobierno, universidades, investigadores independientes, y otras instituciones sin fines de lucro.
- 1.20 *Publicaciones.* Se ha brindado apoyo para mejorar el formato y contenido de las publicaciones de resultados de las encuestas, así como para perfeccionar los modos de difusión de las mismas. En esta línea, los países han publicado una serie de documentos tanto de resultados como metodológicos y han sido difundidos en forma impresa (libros, boletines de prensa, reportes y folletos) y electrónica (CD-Rom e Internet). Al respecto, los países participantes se encuentran desarrollando sus respectivas páginas electrónicas para divulgar sus documentos, varios de los cuales están actualmente disponibles en el sitio de la Unidad de Pobreza y Desigualdad del BID (<http://www.iadb.org/sds/pov>), en donde se encuentran enlaces a las oficinas de estadística que cuentan con páginas electrónicas, así como a la CEPAL y el Banco Mundial.
- 1.21 *Organización de un Banco de Datos.* Se ha conformado un Banco de Datos que contiene 180 bases de datos de encuestas de hogares de 21 países de la región, disponible de modo inmediato a los usuarios internos del Banco, y con proyección de ampliar su acceso a usuarios externos. Este Banco de Datos incluye las más de 80 bases de datos encuestas de hogares que han sido estandarizadas en el formato y documentadas por la CEPAL en función a los antecedentes metodológicos disponibles y al contenido de los archivos.

C. Incorporación de Nuevos Países

- 1.22 De acuerdo al Plan de Operaciones del Programa MECOVI y siguiendo la recomendación del Comité Directivo de incorporar dos (2) nuevos países por año al Programa, a la fecha, se han incorporado Nicaragua (junio 1998), Bolivia (mayo 1999) y Guatemala (septiembre 1999). Se está finalizando el proceso de incorporación de Ecuador, Honduras y República Dominicana. Otros países que han expresado su interés en participar en el Programa MECOVI son Costa Rica, Colombia, Haití, Panamá, y Venezuela. Al respecto, el BID ha ejecutado una cooperación técnica a fin de explorar e identificar las necesidades de asistencia técnica, material y financiera para desarrollar un sistema de encuestas de hogares en los países candidatos a ser incorporados en el programa. Asimismo, dicha cooperación tendrá como objetivo identificar fuentes potenciales de financiamiento y desarrollar un plan de acción para hacer efectiva la incorporación de nuevos países al programa en el periodo 2002-2004.
- 1.23 Ecuador. En el caso de Ecuador el apoyo a las encuestas de hogares se está dando en el marco de un préstamo solicitado por el Gobierno al BID para financiar las actividades del Censo del 2001 y para fortalecer al Instituto Nacional de Estadística y Censos (INEC).
- 1.24 Honduras. Se han realizado tres misiones exploratorias, la primera del 18 al 21 de septiembre, 2000, la segunda del 14 al 18 de mayo de 2001, y la última del 13 al 16 de agosto de 2001. En conjunto con el Instituto Nacional de Estadística, se elaboró un plan de operaciones. Los recursos para financiar el apoyo al fortalecimiento del Sistema Estadístico Nacional de Honduras provienen de un préstamo del BID complementado con recursos de la cooperación sueca y británica.
- 1.25 República Dominicana. Se llevó a cabo una misión indagatoria del 10 al 13 de octubre, 2000, se ha preparado una primera propuesta de plan de operaciones. El Banco Mundial ha aprobado

recursos iniciales para apoyar el mejoramiento de las encuestas de hogares en el país tanto en el Banco Central como en la Oficina Nacional de Estadística.

- 1.26 Caribe angloparlante. En coordinación con el PNUD, el Banco de Desarrollo del Caribe (CDB) y la agencia de cooperación británica se está trabajando con los países del Caribe inglés para poner en marcha un programa de fortalecimiento institucional, capacitación y asistencia técnica en la sub-región. Este programa estaría modelado en la experiencia del Programa MECOVI en América Latina. Se espera que pueda iniciar operaciones en 2003. Se han tenido dos reuniones de coordinación con los países de la sub-región y se cuenta con compromisos de financiamiento firmes de varios donantes bilaterales y multilaterales.

D. Evaluación

- 1.27 Se llevó a cabo una evaluación de las actividades y resultados del Programa MECOVI en Argentina, Bolivia, El Salvador, Nicaragua, Paraguay y Perú. La finalidad de este ejercicio es obtener recomendaciones que permitan mejorar la implementación del Programa en dichos países, y obtener lecciones aplicables en aquellos que se han incorporado al MECOVI más recientemente.

E. Coordinación

- 1.28 El Comité Directivo es la instancia encargada de dictar el rumbo del Programa, está compuesto por el Vicepresidente del BID, el Vicepresidente para América Latina y el Caribe del Banco Mundial, y el Secretario Ejecutivo de la CEPAL. A la fecha ha tenido tres reuniones, el 20 de febrero de 1997, el 27 de julio de 1998, y el 17 de marzo de 2001.
- 1.29 Durante la última reunión del Comité Directivo, 17 de marzo de 2001 en la sede de la CEPAL en Santiago, los miembros del Comité coincidieron en subrayar la importancia del trabajo que el Programa ha realizado en sus primeros cuatro años de labores. Además, se comprometieron a seguir apoyando los trabajos de coordinación y monitoreo en las tres instituciones copatrocinadoras, y a continuar la búsqueda de recursos para apoyar a los nuevos países que se incorporen al MECOVI y para seguir con las actividades regionales del Programa.
- 1.30 El Consejo Asesor está conformado por expertos en el tema de encuesta de hogares y tiene como función dar consejos técnicos para la ejecución del Programa. A la fecha se ha reunido en dos ocasiones, 17 de enero de 1997 y el 6 de noviembre de 1998.
- 1.31 El Comité de Coordinación está formado por los coordinadores del Programa MECOVI en el BID, el Banco Mundial y la CEPAL, tiene como función la ejecución de las actividades básicas del Programa.

F. Conferencia Estadística de las Américas (CEA)

- 1.32 El Programa MECOVI formó parte de la agenda de la primera reunión de la Conferencia Estadística de las Américas (9-11 de mayo de 2001, en la sede de CEPAL en Santiago), y sus actividades quedaron incluidas en el programa de trabajo estadístico internacional para América y el Caribe 2001-2002.

Cuadro 1. Programa MECOVI: Encuestas de Hogares.

Módulos	Argentina (a)		El Salvador (b)		Perú (c) (98-01)				Paraguay (d)		Nicaragua (e)		Bolivia (f)	Guatemala (g)	
			98 (h)	99 (m)	00	I	II	III	IV	97/98 (h)	99 (h)	00/01 (l)	98 y 01 (h)	99 (k)	99 IV (h)
1. Características del hogar y sus miembros															
2. Migración															
3. Vivienda															
4. Salud (general)															
5. Salud de la mujer															
6. Antropometría															
7. Educación															
8. Participación de padres en educación															
9. Acceso a programas sociales															
10. Empleo e ingreso															
11. Actividad independiente del hogar															
12. Actividad agropecuaria															
13. Gastos del hogar															
14. Ahorro financiero, compras a crédito															
15. Uso del tiempo															
16. Remesas															
17. Seguridad ciudadana															
18. Choques adversos															
19. Etnia															
20. Participación ciudadana															

NOTAS:

(a) Encuesta Permanente de Hogares (EPH). Esta encuesta se encuentra todavía en la fase de reformulación dentro del contexto del Programa MECOVI.

Los módulos señalados son los que se implementan como parte de la EPH tradicional en mayo y octubre de cada año.

(b) Encuestas de Hogares de Propósitos Múltiples (EHPM).

(c) Encuesta Nacional de Hogares (ENAHO).

(d) Encuesta Integrada de Hogares (EIH) 1997/98 y Encuesta Permanente de Hogares (EPH) 1999.

(e) Encuesta Nacional de Hogares sobre Medición de Niveles de Vida (EMNV).

(f) Encuesta Continua de Hogares (ECH).

(g) Encuesta Nacional de Condiciones de Vida (ENCOVI).

(h) Bases de datos disponibles para usuarios.

(i) Encuestas en ejecución.

(k) Esta encuesta es una versión modificada de la EMNV 1998, la sección de choques consiste en una serie de preguntas diseminadas a lo largo de los módulos del cuestionario enfocadas al impacto del Huracán Mitch. Esta encuesta sólo se ejecutó en las áreas afectadas por el Huracán Mitch.

(m) En proceso de revisión y consistencia de las bases de datos de las encuestas.

11° TALLER REGIONAL DEL MECOVI
METODOLOGÍAS PARA LA INTEGRACIÓN DE BASES DE DATOS DE
ENCUESTAS DE HOGARES

RESUMEN Y CONCLUSIONES

ÍNDICE

	<u>Página</u>
Introducción.....	23
Discusiones y Debates.....	24

Resultados de los Debates

Introducción

El **undécimo Taller Regional del Programa MECOVI** se realizó en la ciudad de Lima, Perú, del 16 al 18 de julio de 2003. El tema central que orientó los debates fue el análisis de las “**Metodologías para la integración de Bases de Datos de las Encuestas de Hogares**”, debido a que cada vez es más frecuente que la toma de decisiones y el diseño de políticas apoye sus acciones haciendo un uso intensivo de los datos que generan las encuestas.

Al igual que en las ediciones anteriores, este undécimo taller se realizó para dar cumplimiento a los compromisos asumidos por la División de Estadística y Proyecciones Económicas de la CEPAL en el convenio establecido con el Banco Interamericano de Desarrollo (BID) y el Banco Mundial en el marco del programa MECOVI.

Es ampliamente reconocido que el avance de la tecnología informática está cambiando los métodos de trabajo e incidiendo de manera positiva en los procesos de producción y explotación de la información estadística, y de manera especial en la que se genera a partir de las encuestas de hogares.

Con esta idea en mente, se decidió congregarse a los responsables del diseño de los sistemas de procesamiento y validación de datos, así como al personal encargado del armado de las bases de datos que se ponen a disposición de los usuarios, con el ánimo de conocer los fundamentos técnicos de sus procedimientos de trabajo, así como para debatir e intercambiar puntos de vista respecto al uso que están haciendo de las herramientas computacionales existentes, las adaptaciones que han realizado y las lecciones que pueden compartir.

El **objetivo general** del taller se orientó a conocer las metodologías que se siguen en los Institutos de Estadística de la región para encarar las tareas asociadas a la integración de las bases de datos que contienen la información generada por las encuestas de hogares, así como las políticas institucionales que se han definido para garantizar el acceso de los usuarios a la información.

Asimismo, como **objetivos específicos** se plantearon los siguientes:

- i) Profundizar en las metodologías y los aspectos técnicos que se utilizan para el diseño de los sistemas de **captura de datos**.
- ii) Evaluar los procedimientos técnicos que se están aplicando para garantizar la **consistencia y confiabilidad** de la información que se recopila.
- iii) Profundizar en el conocimiento de las herramientas computacionales que se utilizan en la región para sustentar los procesos de **integración de variables y generación de tabulados**.
- iv) Analizar las experiencias nacionales que se siguen en materia de **documentación** de procesos.
- v) Debatir acerca de las experiencias nacionales en el uso de los **sistemas integrados** para el procesamiento de información estadística.
- vi) Conocer los detalles técnicos asociados a las características y **arquitectura de los sistemas de base de datos** que facilitan el acceso a la información.
- vii) Conocer las experiencias nacionales en el uso de paquetes estadísticos y debatir acerca de sus bondades y limitaciones respecto a sus capacidades para apoyar el procesamiento y la explotación de la información estadística.

Discusiones y Debates

La participación de los países se orientó a exponer y evaluar los procedimientos técnicos que se siguen en las oficinas nacionales de estadística para atender las distintas fases que se involucran en el procesamiento de una encuesta de hogares.

Se analizaron las metodologías que se han diseñado y se profundizó en el conocimiento de los sistemas de cómputo y las aplicaciones desarrolladas para la captura, validación y el procesamiento de datos así como aquellos que se usan para la explotación de la información, con especial énfasis en los que se orientan a la generación de tabulados y la construcción de bases de datos que faciliten el acceso remoto a los usuarios externos.

En lo que hace relación a la **digitación de la información** de base contenida en los cuestionarios, se informó que las estrategias que se han implementado en los países para atender esta etapa del proceso son de diferente naturaleza y no admiten una única clasificación. La falta de recursos y de estándares adecuados han obligado a desplegar distintas opciones técnicas que no necesariamente se pueden considerar como prácticas óptimas.

Los participantes precisaron que para encarar esta etapa hacen uso de desarrollos propios elaborados en Clíper o en manejadores de bases de datos como Oracle, Informix, entre otras mencionadas. No obstante, también algunos países señalaron los beneficios que les ha reportado hacer uso intensivo de algunos paquetes de cómputo, así como de los sistemas integrados que están disponibles y han sido diseñados para el procesamiento de censos y encuestas, los cuales permiten a los usuarios efectuar todos los procesos con la misma plataforma informática y un ambiente de trabajo único.

Durante las discusiones técnicas se analizaron las bondades y restricciones que ofrece cada aplicación, y algunos participantes indicaron que los lenguajes y paquetes que están en uso en sus instituciones se asocian, en muchos casos, con las restricciones presupuestarias que enfrentan, y desafortunadamente no siempre se justifica su uso desde un punto de vista técnico o por ofrecer las mejores soluciones a los problemas que se derivan del procesamiento de una encuesta de hogares.

En ese sentido, se señaló que los sistemas integrados que están en uso en las oficinas de estadística se han adoptado por las ventajas técnicas que ofrecen, y además porque son de acceso libre y no les representan a los países ningún costo. No obstante, también se indicó que los usuarios enfrentan restricciones respecto al uso adecuado de estas aplicaciones, ya que el aprendizaje que han hecho los técnicos nacionales encargados de procesar las encuestas es totalmente empírico, lo cual en muchos casos se traduce en la subutilización de las capacidades de los lenguajes de programación y los sistemas existentes.

De las discusiones se puede advertir que no existe un estándar que permita ubicar a un paquete o sistema como la herramienta por excelencia para encarar la etapa de captura y validación. En ese sentido, además de las aplicaciones que han sido mencionadas –lenguajes de programación y manejadores de bases de datos–, los participantes identificaron el sistema ISA, CsPro, los paquetes estadísticos Spss y el Sas como las herramientas que mayor uso tienen, sin dejar de mencionar que el sistema IMPS continúa teniendo vigencia a pesar de que este ha sido reemplazado por el Buró del Censo de los Estados Unidos por una herramienta más poderosa y eficiente (CsPro).

Con respecto a la participación de las instancias regionales en la planeación y ejecución de las distintas fases de la encuesta, se analizaron experiencias de aquellos países que apoyándose en la estructura geográfica de las oficinas de estadística han logrado desconcentrar distintas etapas. Además que el trabajo de campo se ha **descentralizado**, se indicó que las oficinas regionales son las responsables de atender los procesos de digitación, validación y consistencia de los datos, y son también las responsables de velar por la calidad de la información recopilada. Al concluir con esta tarea, los archivos electrónicos son enviados a las oficinas centrales para su integración y validación final.

Por otra parte, también se informó que en aquellos países donde la captura, validación y consistencia se realiza centralmente, se han diseñado procesos en línea de tal manera que los cuestionarios siguen una ruta crítica que permite detectar errores y efectuar correcciones, al tiempo que se valida la información y se generan cifras para evaluar su consistencia.

Pocos son los países que aprovechan la presencia de los entrevistadores en el terreno para realizar la digitación y validación de los datos en campo. Las experiencias nacionales que han logrado poner en práctica esta modalidad, hacen uso de computadoras portátiles que contienen la imagen del cuestionario que se utiliza para recabar la información, así como programas especiales para validar la consistencia de la información cuando esta se digita.

Toda vez que los encuestadores han recabado la información, los datos son introducidos a la computadora, haciendo uso de aplicaciones interactivas que permiten efectuar la consistencia en línea de los datos. Las inconsistencias encontradas en la información recopilada son discutidas con los entrevistadores, y en caso de que se considere necesario se efectúan retornos a los hogares para realizar aclaraciones o superar omisiones en la información.

Se discutió que este tipo de prácticas se considera muy atractiva, pero no siempre resultan adecuadas por su costo y por las condiciones de trabajo de las áreas que se visitan. Por ejemplo, en muchos países además de que las zonas rurales tienen problemas de acceso, no siempre ofrecen las condiciones apropiadas en servicios básicos, por lo que la falta de electricidad puede ser uno de los factores que conspiran en contra de adoptar el uso de computadores en terreno. Además, se señaló que en las áreas urbanas la delincuencia se ha extendido, por lo que equipar a los encuestadores con computadoras portátiles podría poner en riesgo su integridad física.

Hubo amplio consenso en afirmar que la **consistencia de la información** constituye una etapa fundamental para garantizar la calidad de los datos que reportará la encuesta. En ese sentido, las prácticas nacionales que fueron analizadas indican que en todos los países se hacen importantes esfuerzos por diseñar programas de consistencia y validación, que reduzcan al máximo los errores humanos y permitan detectar las inconsistencias que pudieran surgir en el registro en campo y digitación de la información proporcionada por los hogares encuestados.

Hubo consenso en señalar que durante la etapa de captura la mayoría de los países realiza una primera validación de la información, a partir de sistemas inteligentes que permiten detectar inconsistencias al momento de digitar los datos originales. No obstante, y a pesar de que en esta fase es posible identificar errores, se indicó que en la medida que la información se encuentra en archivos magnéticos se corren programas especialmente diseñados para identificar omisiones, inconsistencias y valores aberrantes.

La aplicación de algoritmos especializados es recurrente en la medida que los reportes que se generan continúen señalando la existencia de problemas en la información que se está analizando. De acuerdo a lo expuesto por los participantes, no existe una metodología única, ya que mientras algunos países han adoptado la práctica de aplicar los procedimientos de consistencia y validación para la base de datos en su conjunto, en otros casos se reporta la conformación de archivo temáticos que permiten la aplicación de algoritmos específicos para identificar el comportamiento de variables que se consideran claves para evaluar si la evolución del fenómeno en estudio se comporta conforme a lo esperado.

En ese sentido, algunos países acostumbran a generar reportes que contienen distribuciones de frecuencias de las variables más importantes, así como la emisión de listados en los que se identifican los números de folio de los cuestionarios que presentan errores, en tanto que otras experiencias nacionales indican que además de lo anterior han optado por utilizar la información parcial que se va generando, para construir los indicadores del mercado de trabajo, variables demográficas o el ingreso del hogar, con el objetivo de tener sensibilidad acerca del comportamiento y nivel de confiabilidad de los resultados finales.

Resultó de gran interés discutir acerca de la manera en que los países están atendiendo el tema de la **codificación** de las distintas variables contenidas en el cuestionario de las encuestas de hogares. En este

sentido, las opiniones fueron diversas y no necesariamente coincidentes. Si bien todos los participantes reconocieron la importancia de reducir al máximo los procedimientos de codificación manual, quedó de manifiesto lo difícil que resulta abandonar esta práctica, sobre todo en lo que se refiere a variables asociadas a la actividad económica de los miembros del hogar, y de manera especial en lo que se corresponde con su ocupación y la rama de actividad a la cual se dedica la empresa en donde desarrollan sus actividades económicas.

Se mencionó que algunas oficinas nacionales han tratado de sistematizar esta etapa, pero también se hicieron explícitas las dificultades encontradas para asumir esta práctica. El hecho que algunos países cuenten con clasificadores muy detallados, les dificulta que este proceso se asuma en forma totalmente automatizada. Muchas veces las descripciones dadas por los informantes no son muy claras, lo cual puede conducir a errores de clasificación y en ocasiones podría desvirtuar la naturaleza de la información recabada, y por ende generar distorsiones acerca del comportamiento de alguna categoría ocupacional específica.

Se concluyó que en tanto no se cuente con la garantía que esta etapa del proceso se puede realizar de manera automática, es deseable que los países continúen efectuando verificaciones manuales que les permitan asegurar que la clasificación de la ocupación de los individuos se haga conforme a las prácticas convencionales, en el entendido que se sabe que es necesario continuar perfeccionando los clasificadores que se utilizan conforme a las experiencias nacionales, e invirtiendo en acciones de capacitación que permitan la actualización permanente del personal encargado de realizar esta tarea.

La falta de datos o las inconsistencias de la información, es un hecho al que se enfrenta de manera cotidiana el personal que procesa una encuesta de hogares. Las experiencias narradas por los participantes fueron variadas y no admiten una única tipología. De hecho, no se reporta ninguna regularidad entre los países, ya que cada encuesta es diferente e incluso al interior de un mismo país el grado de omisión o error es muy variable, por lo que las decisiones que se deben asumir respecto a la **imputación de datos faltantes o inconsistentes** no se pueden sistematizar.

Es un hecho que todos los países se enfrentan a la necesidad de suplir datos faltantes o a modificar, a partir de criterios técnicos y de la información del cuestionario, registros que reportan inconsistencias. Al respecto se señaló que la falta de respuesta parcial significa adoptar decisiones técnicas con la intención de aprovechar al máximo la información recopilada, y contribuir con ello a reducir la tasa de no-respuesta, en tanto que las inconsistencias se corrigen con el objetivo de asegurar la confiabilidad de la información reportada por los miembros del hogar.

A pesar de que los participantes coincidieron en señalar que las decisiones se van asumiendo en la medida que los errores van apareciendo, también se indicó que para suplir datos faltantes en ocasiones se hace uso del denominado método "hot deck", el cual sustenta su aplicación en algoritmos que han sido ampliamente probados y han demostrado que generan resultados confiables que en ningún caso desvirtúa la naturaleza de los datos originales.

De las opiniones de los técnicos nacionales se advierte la necesidad de insistir en el hecho que las oficinas nacionales de estadísticas deben adoptar como práctica habitual la documentación de todas las etapas asociadas al diseño, levantamiento y procesamiento de las encuestas de hogares, y de manera especial lo que concierne a la fase de imputación de datos y modificación de la información original.

No es común que los usuarios de la información conozcan los detalles asociados a estas prácticas, ni tampoco la documentación que se entrega señale la magnitud de los errores encontrados en la transcripción y validación de los datos, o las decisiones y criterios que se adoptaron para subsanar las inconsistencias encontradas. Sería deseable y aumentaría la credibilidad de las oficinas nacionales, el que se informe acerca de la manera en que los datos fueron corregidos, la justificación de este hecho y los criterios adoptados para efectuar las modificaciones a la base de datos original.

Por otra parte, se indicó que cada vez es más común que las oficinas de estadística generen **variables agregadas** y construyan indicadores que son incorporados a las bases de datos que se difunden. En ese sentido, se informó que esta práctica responde a necesidades expresadas por los usuarios, lo cual agrega valor al producto de la oficina de estadística. Se señaló que, en la medida de lo posible, los métodos de agregación y cálculo se anexan a la documentación que se distribuye, aunque también se reconoció que es necesario mejorar la difusión de estos procesos incluyendo en los informes metodológicos los aspectos técnicos que sustentan estas prácticas.

A este respecto, se indicó la conveniencia de que se hagan explícitos los criterios que permiten clasificar a las observaciones en las diferentes categorías. Por ejemplo, en el caso de las encuestas de hogares es fundamental saber los distintos filtros que se aplican para identificar a los desocupados, activos, inactivos, subempleados, etc., o, cuando se trata de variables asociadas al bienestar, conocer los umbrales que permiten identificar a una familia en situación de pobreza.

Es de gran utilidad señalar con claridad los períodos de referencia que se utilizan, así como incorporar en la documentación los coeficientes que permiten normalizar la información a un período de tiempo específico cuando se trata de variables asociadas a flujos monetarios.

Un aspecto que fue discutido con amplitud es el que se refiere a la **arquitectura de los archivos de datos**, y a las capacidades de los sistemas integrados y los paquetes estadísticos para su manejo. Se hizo mención explícita a la forma jerárquica en que están estructurados los archivos de las encuestas de hogares, y se indicó que sistemas integrados como el IMPS, ISA y CsPro, tienen capacidad de manejar distintos niveles de información en una encuesta de hogares.

En ese mismo sentido, se mencionó que paquetes como SAS, SPSS y STATA, entre otros, aún no poseen esta capacidad y por construcción sólo permiten operar con un archivo a la vez; en tanto que los lenguajes de programación, como dBase, FoxPro, Visual Basic, SQL y Clipper, tienen a su favor la manera en que trabajan con matrices, además que cualquiera de ellos puede trabajar con dos o más archivos al mismo tiempo, ya sea para efectuar cálculos y consultas o para retirar información útil para otros procesos.

Como un aspecto relevante de los paquetes estadísticos se señaló la ventaja de disponer de macro instrucciones, vale decir comandos que permiten realizar una gran cantidad de cálculos que en un lenguaje de programación como los mencionados, o en un manejador de base de datos, requeriría de decenas de líneas de código para programar el algoritmo de cálculo. Además, estos paquetes son superiores también en cuanto herramientas para el análisis descriptivo o secundario de los datos, en virtud de su mayor nivel de especialización.

Una de las etapas a la cual hay que dedicarle mucho tiempo de programación es la que se refiere a la **generación de los tabulados**. Para atender los requerimientos de los usuarios, los países hacen uso de diversas herramientas informáticas. Existen casos en los que se recurre a programación *ad-hoc* para efectuar los cálculos, y posteriormente los formatos de salida son elaborados en Excel, otros que utilizan las capacidades de paquetes como SAS y SPSS, o bien hacen uso intensivo de los módulos de tabulación de los paquetes integrados.

En todos los casos, el esfuerzo dedicado a esta tarea demanda de parte de los responsables del desarrollo de sistemas el contar con herramientas flexibles, ya que atender los requerimientos especiales de información ocupa una parte importante del tiempo de trabajo de los funcionarios encargados de administrar las bases de datos y de atender a los usuarios especializados.

A pesar de las facilidades que ofrecen el SAS y el SPSS para generar cruces de variables y efectuar cálculos adicionales, se indicó que subsiste el problema de disponer de un formato adecuado para difundir los resultados de las encuestas, y de ahí la necesidad de utilizar herramientas como Excel para mejorar la presentación de las tablas de salida. A pesar que el manejo de esta aplicación es de uso generalizado, se señaló que esta tarea insume mucho tiempo y puede representar un cuello de botella para mejorar la oportunidad en la entrega de resultados.

Al respecto, se insistió en las bondades del IMPS, ISA y CsPro, indicando que sus módulos de tabulación son muy apropiados para atender los requerimientos estándar de información, y que en caso de aplicaciones más específicas han demostrado ser una herramienta que está en posibilidad de efectuar cuadros de salida para cruces de distintas dimensiones y utilizar la estructura jerárquica de los archivos manejados con estos sistemas.

En el entendido que es imposible realizar un detalle pormenorizado y fiel de los debates, y a riesgo de que algún aspecto relevante de la reunión no sea consignado en esta relatoría, el grupo del taller identificó un conjunto básico de conclusiones de las cuales interesa dejar constancia.

1. Se reconoce que la región dispone de un importante patrimonio en materia de recursos humanos, con formación y vocación para el procesamiento de encuestas de hogares. Este personal se ha formado, en su mayoría, de manera empírica y con muy pocos apoyos de capacitación formal, situación que debe ser valorada y reconocida debidamente.
2. Las restricciones económicas que se enfrentan en los países no son ajenas a los institutos de estadística, lo cual ha limitado las posibilidades de realizar las inversiones en "software" especializado que se requieren para mejorar las aplicaciones orientadas al procesamiento de encuestas de hogares.
3. Las plataformas disponibles en los países son muy variadas, y el espectro comprende el uso de lenguajes de programación básicos de alto nivel, sistemas integrados de acceso gratuito, manejadores de bases de datos que no necesariamente se adaptan a la estructura de los archivos que generan las encuestas de hogares, y el uso de paquetes estadísticos de fácil manejo pero con restricciones respecto a las necesidades de las oficinas de estadística.
4. En los países se hace un uso intensivo de lo que existe, lo cual ha permitido alentar el ingenio y creatividad de los técnicos nacionales, expresada en el diseño y desarrollo de utilerías propias que permiten acelerar los procesos de trabajo.
5. El aumento persistente de las demandas de información ha contribuido a que los países desarrollen una gran versatilidad para encarar esta tarea, situación que les confiere un manejo y dominio amplio de lenguajes, sistemas y paquetes estadísticos de mayor uso.
6. Para atender las etapas de captura, validación y consistencia de la información se observó la predominancia de lenguajes como Fox Pro, Visual Basic, dbase y de sistemas integrados como IMPS, ISA y CsPro, haciendo énfasis en que estos últimos son de acceso libre y que su diseño está orientado al procesamiento de encuestas.
7. Para encarar la etapa de análisis de datos se destacó la utilidad que representa disponer del SAS, SPSS y STATA. Debido a que estas aplicaciones están orientadas a usuarios especializados, las macroinstrucciones que contienen facilitan su uso y reducen el tiempo de programación, y en muchos casos lo eliminan por completo.
8. A lo largo de las sesiones de trabajo se reiteró la necesidad de distinguir entre las herramientas que utilizan los programadores y los analistas de sistemas para desarrollar códigos, de aquellas aplicaciones orientadas al usuario final. Estas últimas tienen como objetivo facilitar la comunicación entre el usuario y el sistema, a partir del uso de ventanas auto explicativas que facilitan el diálogo. A este respecto, también se manifestó que para hacer un uso óptimo de este tipo de herramientas es necesario que los usuarios tengan un conocimiento adecuado de los procedimientos matemáticos y estadísticos que generan resultados, con el propósito de hacer un uso eficiente de los paquetes y aprovechar al máximo las ventajas que estos ofrecen.
9. La diseminación de las bases de datos concitó la atención de los participantes, y se insistió en la importancia de conocer los distintos formatos existentes que facilitan el acceso de los usuarios a la información de las encuestas de hogares. A este respecto, se señaló que hay aplicaciones

comerciales que incorporan algoritmos que permiten traducir una base de datos a los más variados formatos, como dbf, xls, sav, txt, com, sav, sas, stata, informix, solaris, acces, oracle, entre muchos otros, con la posibilidad de hacer adaptaciones para distintas plataformas como windows, dos, mac, sun, unix, apple, etc.

10. Existió amplia coincidencia en la necesidad de continuar trabajando y realizando esfuerzos para mejorar las aplicaciones actuales, con el propósito de hacer más eficientes los procesos, reducir los tiempos de proceso y mejorar los plazos de entrega de resultados, así como la calidad de la información que se produce.

**PROCESAMIENTO DE DATOS DE LA ENCUESTA
PERMANENTE DE HOGARES**

INSTITUTO NACIONAL DE ESTADÍSTICA Y CENSOS (INDEC)

ARGENTINA

ÍNDICE

	<u>Página</u>
1. Encuesta Permanente de Hogares	33
Síntesis del estado actual, mayo 2003	33
Actividades previstas	33
2. Antecedentes Informáticos	34
EPH habitual	36
EPH continua.....	39

1.- Encuesta Permanente de Hogares

Síntesis del estado actual, Mayo 2003

La Encuesta Permanente de Hogares (EPH) es un programa nacional de producción sistemática y permanente de indicadores sociales, que permite conocer las características sociodemográficas y socioeconómicas de la población.

En su modalidad original, se ha venido aplicando en Argentina desde 1973, dos veces al año (mayo y octubre), llegando actualmente, en un plan de incorporación progresiva, a cubrir 31 aglomerados urbanos y un área urbano-rural.

En esta modalidad puntual, la EPH se aplica en el marco del esquema de centralización normativa y descentralización ejecutiva, siendo realizados los relevamientos por las Direcciones Provinciales de Estadística bajo las normas técnicas y metodológicas fijadas y monitoreadas por el equipo central de la EPH en INDEC.

En base a esta encuesta se proporcionan regularmente, entre otros resultados, las tasas oficiales de empleo, desocupación, subocupación y pobreza, que se complementan con la producción habitual de más de 2.500 tabulados anuales, bases de datos y publicaciones.

También se realizan regularmente tareas de desarrollo de metodologías y de análisis de resultados.

La Reformulación de la EPH se ha desarrollado con el objetivo de reelaborar la metodología de medición y las formas de operación atendiendo a las características socioeconómicas actuales, a las nuevas modalidades de inserción en el mercado de trabajo y a su dinámica de cambio.

La Reformulación de la EPH abarca aspectos temáticos, en función de la adecuación de los instrumentos de captación y las dimensiones de análisis y presentación de resultados; muestrales, tendientes a adecuar el diseño a cambios más frecuentes en los fenómenos medidos; y organizativos, ajustando la modalidad de trabajo a los diseños temáticos y muestrales establecidos.

En función del desarrollo técnico de esos aspectos de la reformulación, la nueva encuesta cuenta con cuestionarios rediseñados, a aplicar en una muestra distribuida en el tiempo, bajo una modalidad de relevamiento continuo y con mayor frecuencia de presentación de resultados.

Ambas modalidades de la EPH (habitual- puntual y reformulada- continua) se encuentran actualmente en campo, relevándose en todo el país (área cubierta), en el esquema metodológico de transición de un modelo a otro.

Actividades previstas

EPH Puntual:

- Relevamiento, en Mayo, en todos los aglomerados urbanos que abarca la encuesta. Incluye cuestionario adicional sobre Plan Jefas/jefes de Hogar, para caracterización y medición de su incidencia sobre desocupación y pobreza. El relevamiento de Mayo está relacionada con el del aglomerado urbano-rural ya realizado (marzo, época de cosecha). Relevamiento en octubre de los 3 aglomerados urbanos de reciente incorporación y en el aglomerado urbano-rural (septiembre, etapa de receso).
- Difusión de Informes de Prensa:
 - ◆ Cinco informes de prensa correspondientes a la onda de mayo de 2003 (Mercado de trabajo total de aglomerados, Mercado de trabajo Gran Buenos Aires, Pobreza total de aglomerados, Pobreza Gran Buenos Aires, Indicadores sociodemográficos y socio-económicos).

- ◆ Tres informes de prensa en relación a la onda de octubre sobre los aglomerados de más reciente incorporación y el aglomerado urbano-rural.
- ◆ Siete informes (junio-diciembre, difusión mensual) de Valorización de la Canasta Básica Alimentaria.
- Elaboración de las restantes herramientas de difusión (bases, tabulaciones, publicaciones, etc.)

EPH Reformulada:

- Relevamiento en el total de los aglomerados urbanos que releva la encuesta (exceptuando los 3 de reciente incorporación), en curso desde Enero 2003, hasta Diciembre en la modalidad continua. Encuesta de nuevo diseño, en campo de manera permanente (distribución de la muestra en el tiempo). Incluye cuestionario adicional sobre Plan Jefas/jefes de Hogar.
- Avances en los desarrollos metodológicos de la Reformulación, y de indicadores especiales (Pobreza y otros).
- Difusión de la información procedente de la EPH reformulada para distintos dominios de estimación.

TRANSICIÓN: Programa de presentaciones, en distintos ámbitos, de la transición de EPH Puntual a Reformulada, continuando contactos con interlocutores especializados de organismos oficiales, instituciones sociales, académicos, expertos nacionales y extranjeros, usuarios calificados de la información, etc. Profundización de las jornadas de capacitación a periodistas de distintos medios, para la comprensión de las características técnicas de la información estadística disponible de la EPH puntual y las implicancias de su Reformulación.

2.- Antecedentes informáticos

Año 1987/88

- En 1987, la Encuesta Permanente de Hogares incorporó el uso de las microcomputadoras al ingreso de datos y al procesamiento de la información. Este hecho permitió por una parte la descentralización del ingreso de datos y al mismo tiempo el desarrollo de herramientas informáticas que mejoraron los controles y el análisis de la información.

Si bien se tendió a la descentralización, se consideró necesario mantener centralizado el desarrollo de aplicaciones para homogeneizar criterios en el ingreso de los datos e implementar los mismos controles en los distintos aglomerados del país.

- Se desarrolló un Sistema de ingreso (en clipper) el cual permitía:

Ingresar la identificación de las viviendas seleccionadas (a fin de efectuar los controles de diseño)

Ingresar los cuestionarios de la EPH

Efectuar la expansión la muestra.

Obtener salidas de control de la información ingresada.

En lo que respecta al ingreso de los cuestionarios, el sistema tenía incorporado la secuencia de las preguntas, el aceptar sólo los códigos válidos (precodificados) y algunas consistencias básicas.

Sobre dicho sistema se fueron efectuando mejoras tendientes a su adaptación a las necesidades de la EPH. Dichas mejoras se relacionan con:

Nuevas consistencias
Nuevos listados
Distintos métodos de selección de viviendas
Cambio del cuestionario básico de la EPH.

Año 1989

- Se comenzó la transferencia (instalación y capacitación) del Sistema de ingreso de datos a las Direcciones Provinciales de Estadística (DPE).

Año 1990

- En el ingreso de los cuestionarios se tendió a lo que denominamos “Ingreso Directo ó Simultáneo”, es decir, que se ingresaran las encuestas durante el operativo de campo, para que el encuestador pueda volver a tomar contacto con el entrevistado ante incompletitudes e inconsistencias.

El “Ingreso Directo” requirió también ajustar el sistema a los efectos de contar con la mayor cantidad de consistencias y validaciones on- line.

Año 1992

- Se realizó una prueba piloto de la reformulación, para lo cual se desarrolló una aplicación con el paquete ISSA (Integrated System for Survey Analysis).

Año 1996

- Se encaro un cambio en el sistema de ingreso con el objetivo de que se ajustara fácilmente a los cambios de cuestionario y efectuara una mayor cantidad de controles en cuanto a calidad de la información.

Esto se logró parametrizando en bases de datos las preguntas de los cuestionarios, y toda la información referente a validaciones, saltos, consistencias etc. de manera que el agregar, sacar ó modificar una pregunta y/o una consistencia no implicara modificar la estructura del sistema, sino simplemente modificar ó agregar un registro a una base de datos.

Otras características del sistema es poder contar con planillas, reportes y procesos que colaboran con los controles de gabinete, el ingreso de las encuestas y la depuración de la información.

Para el comunicado de prensa, se aprovechó la modalidad del ingreso simultáneo para obtener a través del sistema los datos necesarios para la elaboración de las tasas que se informan en el mismo; siguiendo las DPE con la etapa de última depuración.

- Se llevó a cabo la segunda prueba de la reformulación de la EPH, en INFORMIX , como previsión al sistema que se implementaría en la DPE.

Año 1997

- Se desarrolló un Sistema de control de calidad, mediante el cual se buscó disponer en médios magnéticos de información proveniente de distintas fuentes : planillas que completaban en campo los encuestadores, supervisores y recuperadores; listados del personal; bases de datos con la identificación de las direcciones seleccionadas; base de datos con los datos de las encuestas ingresadas, para permitir distintas líneas de análisis, entre ellas:

Colaborar con la organización del trabajo de campo para mejorar la calidad de los datos

Disponer de datos acerca del desempeño de los encuestadores en campo y brindar a los mismos un detalle de su trabajo.

Detectar porcentajes de cambios producidas en la supervisión, aportando datos para la elaboración de estrategias de capacitación a los encuestadores.

Año 1998:

- Se realizó una prueba piloto utilizando notebook. El objetivo de esta prueba fue observar y analizar la influencia del uso de las computadoras portátiles en la situación de entrevista. Se utilizó un cuestionario reducido y se adaptó el sistema de ingreso actual de la EPH de forma de poder utilizarlo para realizar la entrevista en campo.

EPH habitual

Sistema Actual de Captura, Procesamiento y Explotación

El sistema de gestión y procesamiento abarca desde la incorporación de las viviendas seleccionadas del MMNU hasta la obtención de una base consolidada.

El sistema utilizado por la EPH se basa en el registro de los datos en los cuestionarios impresos en papel.

La EPH habitual cuenta con un sistema desarrollado íntegramente en clipper (1985) que permite los siguientes procesos:

Preparar el operativo

- ***Incorporar al sistema los archivos provistos por el Marco Muestral***

Actualizar la muestra, obteniendo las direcciones para las etiquetas, cartas, hojas de ruta y los códigos de viviendas válidas.

Incorporar al sistema los archivos en el formato que los recibimos del Marco Muestral, realizar controles sobre esos datos. (txt)

- ***Módulo de envío de archivo a las pcias.***

Se envían a las provincias los insumos para empezar el ingreso de la información

-Instructivos

-Ejecutables

-Bases (area.dbf, area_hog.dbf, hogar.dbf, personas.dbf, perso2.dbf)

-Archivos del MMNU

Proceso de Captura y Consistencias

El sistema de ingreso y procesamiento está descentralizado. Todos los procesos que se detallan son realizados en las Direcciones Provinciales de Estadística.

La EPH cuenta con un sistema de ingreso íntegramente desarrollado en Clipper que permite:

- ***Ingresar al sistema los cuestionarios***
- ***Ingreso de otros materiales de campo***

Posibilidad de ingresar, planillas que se utilizan en el trabajo de campo Y de todos los materiales que permiten consistir y evaluar el trabajo de campo.

- **Posibilidad de correr otros procesos de consistencias Batch y otros controles**

Control de Completitud (relación entre las viviendas seleccionadas y las ingresadas)

Impresión de Planillas R1 con la información de los cuestionarios ingresados

- **Realizar consistencias en línea**

El sistema cuenta con una base de consistencias que utiliza para validar la información al momento del ingreso del cuestionario.

- **Controles de Incompletos, repetidos**

- **Variabkes construidas definidas en programa de captura**

Condición de actividad

Categoría ocupacional

- **Variables Codificadas**

En la Eph Habitual la rama de actividad y la Ocupación son codificadas previamente y validadas por el sistema al momento del ingreso.

- **Calculo de tasas**

Una vez finalizado el ingreso y realizado los controles y consistencias las bases son remitidas a EPH-Indec (hogar.dbf , personas.dbf, perso2.dbf, area.dbf, area_hog.dbf)

Los procesos que se realizan a las bases recibidas son

En EPH- Indec se continúan con los siguientes procesos

- **Imputación (solo aglomerado: Gran Buenos Aires)**

-Método Hot-Deck: se utiliza en casos de ausencia o rechazo del cuestionario individual, personas de las que disponemos de los datos demográficos. Se imputa condición de actividad, categoría ocupacional y nivel educativo.

-Hogares que fueron respuesta en la onda anterior y que en la onda actual existe alguna información faltante demográfica y/o educación (solapamiento).

Se informa a los usuarios la cantidad de casos imputados y los mismos pueden ser identificados ya que existe un campo (imputa) con marca que los identifica.

- **Ponderación:** posibilidad de aplicar la expansión a la muestra y el ajuste proyección demográfica.

(Variable construida: pondera).

- **Deciles :**Posibilidad de distribuir a los hogares y a la población según escalas decilicas.

En este proceso se generan variables construidas de deciles que se detallan en el Documento Base Usuaría Ampliada **Calculo de tasas y desvíos**

Posibilidad de calcular tasas y desvíos: muestrales y/o ponderadas según permisos asociados.

- **Estructuras**

Posibilidad de calcular las estructuras como resultado de los datos ingresados y compararlos con información histórica.

Posibilidad de establecer universos: Ej: Aglomerados, Regiones, Total Aglomerados EPH

Una vez finalizado todos los procesos detallados contamos con una base Total

(Aglomerados - Eph)

Sistema de Explotación – Bases Total_Eph

Se cuenta con sistemas en Clipper que permite generar

- **Comunicados de Prensa**

- Indicadores socioeconómicos para 31 aglomerados

- Mercado de trabajo: principales Indicadores de los aglomerados Urbanos

- Mercado de trabajo: principales indicadores del aglomerado Gran Buenos Aires

- Incidencia de la Pobreza y de la Indigencia en los aglomerados Urbanos

- Incidencia de la pobreza y de la Indigencia en el Gran Buenos Aires

- **Base Usuaría Ampliada**

La bases usuarias (hog_bua.dbf y per_bua.dbf) son bases desagregadas que contienen todas las preguntas del cuestionario, las variables construidas (estado, inghora, imputa, variables generadas en el proceso de deciles).

Se encuentra a disposición de cualquier usuario

Internet <http://www.indec.mecon.gov.ar/>.

Se puede acceder a las bases de cada Aglomerado por separado, y a la del Total de Aglomerados.

- **Documento de Base Usuaría Ampliada**

Contiene información y recomendaciones para su utilización, desde diseño de estructura de las bases, utilización de la Tabla de Errores de Muestreo, variables construidas, etc.

- **Tabulados**

La EPH cuenta con un Tabulador desarrollado en Clipper que nos ha permitido generar tabulados muestrales y ponderados, Tabla de errores de muestreo, frecuencias, identificar aquellos datos cuyo coeficiente de variación es superior al 10%, frecuencias.

- Tabulados Básicos

- Tabulados Complementarios

-Publicaciones Especiales

- **Publicaciones**

- Anuario Estadístico de la República Argentina

- Publicación Encuesta Permanente de Hogares

- **Otros Documentos**

- Diseño de registro de la Base Usuaría Ampliada (BUA) para los Aglomerados

- Diseño de registro de la Base Usuaría Ampliada (BUA) para el Total de Aglomerados

- Utilización de Clasificadores de Ocupación en las BUA
- Novedades de cada Onda
- Clasificador Nacional de Ocupaciones Internet <http://www.indec.mecon.gov.ar/>.

EPH continua

Para la Encuesta continua desde el punto de vista informático se parte de cambios importantes

- Los controles de información y el envío de la información será mensual.
- La información llegará a Indec mensualmente y se consolidará en una base Nacional. Se desarrollaron programas que acompañan la metodología de trabajo.

- ***Programa de Instalación***

Se envían a las DPE programa de instalación que genera el directorio de trabajo con todos los insumos necesario para comenzar el ingreso de la información:

- Base con viviendas seleccionadas (con los archivos de MMNU incorporados)
- Bases (area.dbf, area_hog.dbf, viv_gral.dbf, hogar.dbf, personas.dbf, perso2.dbf, viv_gra.dbf, hog_cara.dbf, viv_cara.dbf, viv_pers.dbf)
- Incorpora la información provenientes de otros trimestres. Las preguntas de características de vivienda y hogar solo se preguntan una sola vez.

Proceso de Captura y Consistencias

El sistema de ingreso y procesamiento está descentralizado. Todos los procesos que se detallan son realizados en las Direcciones Provinciales de Estadística.

La EPH cuenta con un sistema de ingreso íntegramente desarrollado en Clipper que permite:

- ***Ingresar al sistema los cuestionarios***
- ***Ingreso de otros materiales de campo***
Posibilidad de ingresar, planillas que se utilizan en el trabajo de campo y de todos los materiales que permiten consistir y evaluar el trabajo de campo.
- ***Posibilidad de correr otros procesos de consistencias Batch y otros controles***
Control de Completitud (relación entre las viviendas seleccionadas y las ingresadas)
Impresión de Planillas R1 con la información de los cuestionarios ingresados
- ***Consistencias en línea***
Se están incorporando las consistencias para los nuevos cuestionarios (vivienda, hogar, individual) que serán incorporadas en el programa de Ingreso.
- ***Controles de Incompletos, repetidos***
- ***Variables construidas definidas en programa de captura***
Condición de actividad
Categoría ocupacional

- **Variables Codificadas**

La Rama de la Actividad y Tarea de la Ocupación no se codifican en DPE. Se ha comenzado a utilizar Codificación automática y/o asistida en bases trimestrales de Eph-Indec

- **Calculo de tasas**

Envío de la base mensual de la DPE a Eph-Indec

Se ha desarrollado un Sistema que arma archivos con un formato determinado que solo se genera si se han seguido los pasos de controles definidos (consistencias, completitud, repetidos, incompletos)

La base mensual es recibida en Indec-Eph en la primer semana del mes siguiente.

Recepción y Consolidación de los datos

Se cuenta con un sistema en red que tiene definidos roles de usuarios que garantizan que los procesos sean corridos por aquellos que están autorizados.

En EPH- Indec se continúan con los siguientes procesos

- **Levanta los archivos remitidos por la DPE y los incorpora a la base trimestral.**

- **Calculo de la No Respuesta**

Permite procesar información por estrato y por grupo de rotación, comparar con datos históricos.

- **Control y aprobación de la muestra**

Permite controlar y aprobar la muestra comparando con los datos históricos de trimestres anteriores (viviendas seleccionadas, viviendas respuesta, viviendas por causa de no respuesta). Solo se procesa después de haber recibido los tres envíos mensuales.

- **Calculo de tasas (Comunicado de Prensa)**

Solo se procesa después de haber recibido los tres envíos mensuales.

- **Ponderar y ajustar a la Proyección**

Solo se procesa después de haber recibido los tres envíos mensuales.

- **Procesos de deciles**

Solo se procesa después de haber recibido los tres envíos mensuales.

Paralelamente se desarrolla este mismo sistema (**Recepción y Consolidación de los datos**) en Oracle. Este sistema abarca la totalidad de las tareas y permite analizar y visualizar la información ya ingresada desde distintas formas de agrupamiento

Durante un tiempo deberán coexistir los dos sistemas. En este momento estamos controlando los resultados de los procesos generados en ambos sistemas.

**PARTICIPACIÓN DE LA INFORMÁTICA EN EL
DISEÑO, DESARROLLO Y EXPLOTACIÓN DE
ENCUESTAS A HOGARES**

**MARIA TERESA ARRIOLA, GRACIELA CANESSA,
RAQUEL SALVA Y ELBA TONINI**

INSTITUTO NACIONAL DE ESTADÍSTICA Y CENSOS (INDEC)

ARGENTINA

ÍNDICE

	<u>Página</u>
Introducción	43
Sistema de Ingreso ENGH 96/97	43
Integración de la Base de Datos – Transformación – Imputación – Resultados	44
Síntesis Global del Ciclo de Evolución de los Datos.....	45
Base a Usuarios – Descripción.....	46
Desafíos y Proyectos de la Nueva Encuesta en el 2003	46
Conclusión.....	47

Introducción

El procesamiento de datos de la ENGH (Encuesta Nacional de Gastos de los Hogares) comprende un conjunto de tareas que comienza en el desarrollo e implementación del Sistema de Ingreso para acompañar la etapa de captura de datos. Inmediatamente después, y mientras el ingreso de la información está en pleno desarrollo, se origina la necesidad de efectuar los procesos de control de calidad de la información que se está ingresando. Dentro de esta categoría de procesos podemos mencionar el análisis de tasas de no respuesta o los estudios de las cantidades de registros y de valores extremos, pero también, la obtención de estructuras de gastos preliminares. Para poder implementarlas es necesario introducir procedimientos de generación de variables transformadas y de agregado de variables. Un poco más tarde en el ciclo de vida del procesamiento de la encuesta aparece la necesidad de imputación de variables faltantes para evitar el sesgo que se produciría si se presentaran los resultados considerando únicamente los datos brindados por los hogares respondientes. Finalmente, la producción de los cuadros para la publicación de resultados y la generación de la Base a Usuarios constituyen el corolario para el ciclo del desarrollo.

Sistema de Ingreso ENGH 96/97

El Sistema de Ingreso para la ENGH 1996/1997 fue desarrollado en INDEC por el equipo informático de la Dirección de Estudios de Ingresos y Gastos de los Hogares y se instaló en once centros regionales de ingreso a lo largo del país. Cada centro de ingreso incluía una o varias provincias. Operativamente, el centro regional estaba organizado mediante los siguientes puestos de trabajo:

- Un Coordinador Técnico, quien era responsable del buen funcionamiento del servidor y del motor INFORMIX instalado en el centro;
- Un Encargado Regional de Ingreso, quien coordinaba la tarea de los editores y emitía las planillas de control y monitoreo del ingreso y del operativo y
- Un conjunto de Editores, quienes ingresaban, modificaban, consistían y listaban las encuestas, generaban los listados de errores y los balances financieros.

Todos los módulos del sistema de ingreso fueron implementados en INFORMIX-4GL – versión 4.10, así como la generación de la base de datos. El sistema se comunicaba con el usuario a través de menús que daban acceso a pantallas similares a los bloques o páginas de los formularios y guiaba al editor para el ingreso de cada encuesta en forma pautada. La validación en el ingreso de datos se clasificó en tres grandes grupos, dependiendo de la característica de la pregunta:

- Efectuando el chequeo de existencia en una tabla referencial (por ejemplo de artículos).
- Validando que la respuesta estuviera dentro de un conjunto de valores permitidos codificados en forma previa, conocimiento que se incluyó en los programas (por ejemplo valores válidos para categorías ocupacionales o lugares de adquisición).
- Verificando la inclusión del valor ingresado en un intervalo almacenado en una tabla referencial (por ejemplo para la validación del monto, el rango de precio mínimo y máximo por artículo).

Ante la falla en una validación, para los dos primeros casos, el sistema no permitía el ingreso del dato, para el tercer caso se permitía el ingreso, pero el registro se marcaba como inconsistente para ser evaluado y eventualmente corregido con posterioridad.

Para garantizar la consistencia y coherencia global de la encuesta, el sistema implementaba un conjunto de pautas clasificadas en advertencias y errores. Dichas pautas podían estar circunscriptas a un bloque de preguntas de un formulario determinado (por ejemplo, verificar que si existe alumbrado público en la cuadra en la que se encuentra la vivienda, entonces debe existir electricidad) o cruzar información de

diferentes bloques (por ejemplo, si el hogar tiene electricidad entonces debe declarar gasto para el código correspondiente).

El Sistema implementaba mecanismos de seguridad asignando diferentes roles a los usuarios de acuerdo a su puesto y otorgando un conjunto de permisos a cada rol. Así, por ejemplo un editor no podía dar de baja a una encuesta, atribución que sí estaba asignada al Encargado Regional.

La encuesta se desarrolló en conjunto con las veinticuatro Direcciones Provinciales de Estadística. Allí funcionaba un equipo de analistas que codificaba la información contenida en los formularios relevados por los encuestadores y verificaba manualmente un conjunto mínimo de pautas de consistencia y completitud. Los formularios eran enviados luego al centro correspondiente a la provincia para su ingreso. Una vez allí, el encargado regional distribuía las encuestas entre sus editores, administrando planillas de seguimiento de las encuestas. Luego de ingresada la encuesta se emitían los listados de datos ingresados y los listados de errores con las inconsistencias detectadas por el sistema. El listado y los formularios viajaban nuevamente a la provincia para que los analistas determinaran las correcciones y devolvieran el listado corregido al Centro de Ingreso. Los editores registraban las modificaciones, y la encuesta era finalizada.

El sistema emitía un conjunto de planillas semanales (a imprimir en forma anticipada, dado que constituía el material para la salida a campo) y un conjunto de planilla mensuales (a imprimir a la finalización de cada mes, pues contenían información relevante para el control del operativo).

La preparación para la instalación del Sistema incluyó la generación de bases de datos fijos específicos para cada centro, en particular, la tabla correspondiente a la muestra fue distribuida selectivamente de forma que cada centro tuviera acceso únicamente a las viviendas que le correspondieran. Con referencia al material para el operativo, el sistema de ingreso emitía las hojas de ruta para el encuestador y las planillas para las supervisiones de rutina, las cuales se imprimían en forma semanal anticipada, para disponer de las mismas en forma previa a la salida a campo.

Integración de la Base de Datos – Transformación – Imputación – Resultados

Una vez comenzado el ingreso en los centros, la información era enviada semanalmente a INDEC para ser integrada en una única base de datos centralizada. Esto permitía el monitoreo del ingreso y la consolidación de la información a nivel nacional.

Para poder analizar y permitir la posterior explotación de los datos fue necesario desarrollar los procesos de transformación, los cuales incluyeron, entre otros:

- Mensualización de los gastos para unificar los distintos períodos de referencia que la encuesta releva.
- Mensualización de las cantidades y transformación a la unidad de medida estándar definida para cada artículo.
- Generación de variables agregadas a nivel de hogar (cantidad de miembros y de preceptores, información sociodemográfica del jefe del hogar, gasto e ingreso total, quintiles de ingreso, etc.)
- Generación de variables agregadas a nivel de personas (nivel de instrucción, condición de actividad, categoría ocupacional, etc).
- Imputación de ingresos no declarados.

Conforme se iban recibiendo los datos de los centros de ingreso, iba surgiendo la necesidad por parte del Equipo de Metodología y Análisis de efectuar los controles de calidad y de conocer estructuras de gastos y otros cuadros de cruces de variables. Para ello se desarrolló una aplicación a la que denominamos 'Sistema de Biblioteca', que, mediante una interfase implementada en INFORMIX-4GL, ofrecía una

variada gama de opciones, algunas de las cuales ejecutaban módulos implementados también en INFORMIX-4GL y otras, lanzaban la ejecución de procesos en SAS-versión 6.07, dependiendo de la conveniencia de utilizar un producto u otro según la naturaleza del resultado que se quisiera obtener.

Los cuadros de la publicación se conformaron, entonces, con un subconjunto de los cuadros emitidos por el Sistema de Biblioteca. Toda la información transformada y agregada se adicionó para componer la Base ENGH y luego, mediante selección y agrupamiento se definió la Base a Usuarios de la ENGH.

A medida que los requerimientos del Equipo de Metodología y Análisis de la encuesta crecían, el sistema de Biblioteca iba incrementando su funcionalidad, inclusive incluyó una opción que permitía el acceso a una sesión de SAS a los usuarios del equipo que estuvieran interesados en hacerlo. De esta forma, cada usuario podía generar archivos y salidas de SAS en un área de trabajo propia, sin permiso para sobrescribir los archivos SAS que componían la base maestra..

Como no se disponía en el momento de una interfase INFORMIX-SAS, el hecho de que el sistema de Biblioteca trabajara en forma paralela en dos plataformas, introdujo el inconveniente adicional de tener que mantener actualizada una copia con formato SAS de la base maestra, que se encontraba en permanente evolución, alimentada con los backups que llegaban de los centros de ingreso y con las transformaciones que se le practicaban en INDEC.

También se fueron incorporando al Sistema de Biblioteca procesamientos a pedido de usuarios de otras dependencias del INDEC y de otros usuarios externos al INDEC.

Todos los procesos de transformación fueron debidamente documentados y puestos a punto en forma conjunta con el Equipo de Metodología y Análisis de la encuesta. El mecanismo de imputación se aplicó para los ingresos no declarados.

Dado que se observó que las características de los hogares no respondientes a ingresos eran diferentes a las de aquellos hogares que sí lo hicieron, la omisión de la información de los hogares no respondientes sesgaría significativamente los resultados. A fin de evitar este sesgo, se optó por imputar los ingresos.

La imputación se basó en la existencia de una relación, avalada por la teoría económica, entre los gastos de consumo, ingresos y otras características de los hogares, conocida como la función de consumo. El procedimiento de imputación consistió en estimar, vía regresión múltiple, los parámetros de esta relación, utilizando la información recopilada por la Encuesta sobre los gastos de consumo y las características sociodemográficas de los hogares que respondieron ingresos y utilizar luego los parámetros estimados junto con la información sociodemográfica y de gastos de los hogares que no respondieron ingresos, para imputar los ingresos de éstos.

La imputación se llevó a cabo en dos etapas. En una primera etapa se definió al hogar como dominio de análisis para la función de regresión. De aquí se obtuvo como resultado una primera estimación del ingreso real del hogar. Posteriormente, en la segunda etapa, el dominio de análisis se restringió a los perceptores y, teniendo en cuenta el hecho de que la Encuesta relevó ingresos individuales y a nivel de fuente, se estimaron ingresos para los perceptores que no lo habían declarado y se obtuvo la estimación del ingreso total del hogar como la suma de los ingresos (imputados o no) de sus perceptores.

Fueron publicados resultados basados en cada una de las etapas de imputación. En la Base a Usuarios de la ENGH quedan debidamente identificados los hogares cuyo ingreso fue imputado, para que el usuario los discrimine al evaluar los resultados, si lo considera adecuado.

Síntesis Global del Ciclo de Evolución de los Datos

El Sistema de Ingreso de la ENGH capturó, en cada centro de ingreso, porciones regionales de información que se transmitieron a INDEC para conformar una única base de datos a nivel nacional (Base de Datos Original). La misma se integró y consistió en INDEC para conformar una Base de Datos

Maestra. Ésta fue sometida a procesos de transformación que ajustaron los datos para poder explotarlos y a procesos de imputación que generaron los datos faltantes. En forma paralela, el uso del Sistema de Biblioteca auxilió en la toma de decisiones. Mediante mecanismos de selección y agrupamiento se definió la Base a Usuarios ENGH.

Base a Usuarios – Descripción

La Base a Usuarios se conformó con un conjunto de tablas que representaron a aquellas entidades o sujetos sometidos a estudio por la Encuesta. Así la Base está compuesta por tablas de hogar, personas, ingresos, gastos, cantidades, equipamiento y gastos por forma de pago y lugar de adquisición. La tabla de hogar se relaciona a través de su clave con las tablas de personas, gastos, cantidades y equipamiento y la tabla de personas se relaciona a través de la clave del hogar más la identificación del miembro con la tabla de ingresos.

Las tablas se elaboraron a partir de la información suministrada por los formularios más las variables agregadas por los procesos de transformación e imputación. Del formulario 1 (*Características de los hogares*) se extrajeron las variables para formar la tabla de hogar y personas; del formulario 2 (*Gastos Diarios*) y 3 (*Gastos Varios*) aquellas para las tablas de cantidades, gastos por forma de pago y lugar de adquisición; del formulario 3 para la tabla de equipamiento y del formulario 4 (*Ingresos*) las de la tabla de ingresos.

La información se pone a disposición de los usuarios mediante un CD cuyos directorios principales son *Archivos_texto*, *Documentación* y *Tabulados*. Las tablas que mencionamos en el párrafo anterior están almacenadas en archivos de formato texto con delimitador. La decisión de presentar la información de esta forma es justificada por el hecho de tratarse del formato más portable para poder generar bases en cualquier producto del mercado, y, también, compatible con todas las diferentes versiones de un mismo producto a través del tiempo.

La estructura de estos archivos de texto es idéntica a través de las regiones y está descrita en archivos con formato WORD, conjuntamente con una lista de valores para cada variable y su significado.

En el sub-directorio *Documentación / Metodología* se puede consultar la documentación metodológica de toda la Encuesta, accediendo mediante mecanismos de navegación a archivos WORD conectados mediante links. También se puede acceder en forma directa a consultar los tabulados publicados (estructuras de gastos, cuadros de cantidades y de ingresos). Dichos tabulados se encuentran almacenados en el directorio *Tabulados*, clasificados por la región a la que pertenecen. Toda esta documentación metodológica y tabulados están disponibles también en las publicaciones del INDEC para la Encuesta.

Desafíos y Proyectos de la Nueva Encuesta en el 2003

En este momento nos encontramos en la etapa de implementación del Sistema de Ingreso para la prueba piloto de la nueva Encuesta de Gastos. El mismo se analizó y diseñó con herramientas ORACLE-DESIGNER y se está implementando en ORACLE-DEVELOPER. En esta oportunidad, la política será la de un ingreso descentralizado, llevado a cabo en las Direcciones Provinciales de Estadística, a través de un entorno ORACLE Cliente-Servidor. De esta forma, la base de datos es única e instalada en INDEC y desde las provincias se accede a ella en forma directa.

Nuestra experiencia nos fue convenciendo de la necesidad de encarar el desarrollo e implementación de un sistema de ingreso genérico para encuestas a hogares. La cualidad principal que debe caracterizar a un sistema de ingreso genérico es su capacidad para adecuarse a los cambios que se van produciendo en el diseño de los cuestionarios de la misma encuesta (en la transición entre piloto y definitiva, por ejemplo); o aún para poder aplicarse a dos encuestas a hogares distintas. En respuesta a la necesidad que

planteamos, estamos proponiendo un esquema de diseño e implementación que denominamos *Esquema Parametrizado*.

Tanto en el esquema de análisis y diseño tradicional, como en el esquema que proponemos, el equipo de metodología y el informático co-participan en la elaboración de un conjunto de requerimientos para el sistema de ingreso que el equipo informático implementa en un soft.

Sin embargo, sabemos que en todo soft podemos discriminar claramente entre el conocimiento que el mismo debe incluir y los métodos o algoritmos que debe aplicar para funcionar en forma correcta y completa.

Entonces, en el *esquema parametrizado*, se codifica todo el conocimiento para ser almacenado en tablas, y por otro lado, se diseñan e implementan los métodos que resultan así de naturaleza más general por no incluir el conocimiento dentro de sí. La ventaja sobre el esquema tradicional es que, si se identifica en forma correcta la separación entre conocimiento y métodos, los cambios en el diseño de los cuestionarios sólo implicarán actualizar el conocimiento (modificar datos en tablas) sin que estos cambios afecten los métodos (no se corrigen los programas).

El costo que hay que pagar es el de diseñar la estructura de la tablas para almacenar el conocimiento y desarrollar, implementar y mantener un módulo para el ingreso del mismo. También se requiere una mayor capacidad de abstracción en la programación de los módulos, donde las especificaciones expresadas "en duro" son mínimas o nulas. Creemos que las ventajas superan a los costos.

Los cuestionarios, como instrumento de captación de la encuesta quedan plasmados en una base de datos donde podemos identificar claramente dos grupos bien diferenciados:

- Las tablas que reflejan la estructura de los cuestionarios y todo el conocimiento incluido en ellos.
- Las tablas que guardan los datos propiamente dichos relevados por la encuesta.

¿Cómo se expresa en tablas el conocimiento incluido en los cuestionarios?. En forma simplificada, comenzamos representando a la encuesta en una estructura de árbol. La encuesta está compuesta por cuestionarios, los cuestionarios están organizados o divididos en bloques, los bloques contienen preguntas, cada una de ellas posee características propias (un número que la identifica, un texto que la describe, un conjunto de opciones o valores posibles para la respuesta).

Por fortuna, los bloques no son todos distintos sino que pueden ser agrupados en tipos que responden al mismo patrón. Así, en el soft, definimos un módulo para cada tipo de bloque distinto que identificamos en los cuestionarios. Las diferencias que hay entre dos bloques del mismo tipo quedan expresadas únicamente en las tablas que guardan el conocimiento. Sólo la aparición de un patrón nuevo implicará el desarrollo de otro módulo, mientras que, si se agrega a los cuestionarios un bloque correspondiente a un patrón conocido sólo impacta en la actualización de tablas, no en la programación del módulo.

Conclusión

Como vemos, la informática no debe constituir un compartimiento aislado en el desarrollo de las Encuestas a Hogares sino que su participación e integración en todas las etapas es una condición que amplía las probabilidades de éxito del proyecto. Ponemos nuestros esfuerzos en imprimirle a esta integración un mayor compromiso, aún teniendo en cuenta los costos, pues estamos convencidas que redundará en mayores beneficios, tanto para los que estamos involucrados en la generación de este tipo de proyectos como para los usuarios finales de los datos que proveemos.



ENCUESTA DE HOGARES

FREDDY OCHOA CABRERA

INSTITUTO NACIONAL DE ESTADÍSTICA (INE)

BOLIVIA

ÍNDICE

	<u>Página</u>
I. Introducción	51
II. Objetivos	51
III. Estrategias para la Captura de Información en Computador	51
Encuestas de Hogares MECOVI gestión 1999, 2000 y 2001	51
Encuesta de Hogares MECOVI gestión 2002	52
IV. Procedimientos para Evaluar la Consistencia de la Información.....	52
V. Diseños Computacionales de las Bases de Datos	54
VI. Integración de Variables	54
VII. Arquitectura de las Bases de Datos.....	54
VIII. Generación de Tabulados.....	54
IX. Políticas Institucionales de Acceso a las Bases de Datos	55

I. Introducción

La Encuesta Continua de Hogares (Condiciones de vida/MECOVI) es una investigación por muestreo efectuada a hogares particulares. Se la realizó mediante la aplicación de un cuestionario multitemático, que permite el estudio de las condiciones de vida de los hogares y sus diferentes componentes.

La Encuesta es parte del Programa de Mejoramiento de las Encuestas de Medición de las Condiciones de Vida (MECOVI), y tiene como objetivo el fortalecimiento y la institucionalización de las encuestas a hogares que miden condiciones de vida.

La información de la Encuesta Continua de Hogares, será utilizada para la elaboración de un perfil de pobreza de Bolivia, documento base para la formulación, evaluación y seguimiento de políticas y el diseño de programas de acción en el área social, tendientes a la reducción de la pobreza.

El cuestionario de la encuesta esta organizado por secciones y permite la investigación de las características generales, educativas, empleo, salud, gastos, ingresos, vivienda y servicios básicos de los hogares.

II. Objetivos

El objetivo general de la encuesta es obtener información sobre las condiciones de vida de los hogares, referente a variables socioeconómicas y demográficas, necesarias para la formulación, evaluación, seguimiento de políticas y el diseño de programas de acción en el área social.

En forma específica la ECH - MECOVI se propone:

- Producir una Base de Datos completa con información actualizada de variables importantes, determinantes de las condiciones de vida a nivel de viviendas, hogares e individuos.
- Medir oportuna y apropiadamente los niveles de bienestar y pobreza de la población boliviana en función a sus factores determinantes.
- Aportar con información confiable al estudio de los factores determinantes de la pobreza en la población boliviana.

III. Estrategias para la Captura de Información en Computador

Encuestas de Hogares MECOVI gestión 1999, 2000 y 2001

Para las Encuestas de Hogares MECOVI, de las gestiones 1999, 2000 y 2001, se hicieron transcripciones de datos descentralizadas en nueve oficinas regionales.

Descentralizando la transcripción de datos, permitía que la información pueda ser consistenciada en la misma región. De tal forma que si se presenta alguna incoherencia de información, esta se pueda corregir en primera instancia con la participación de los encuestadores o supervisores de campo, o en última instancia retornando a las viviendas informantes.

Esta captura de información en computador se la realizaba casi en forma paralela al Operativo de Campo, aprovechando de esta forma el regreso del Encuestador a la vivienda en caso que sea necesario.

La información transcrita se agrupa por Unidades Primarias de Muestreo (UPM's), es decir todos los cuestionarios que correspondían a una misma UPM eran almacenadas en un archivo de datos. Una vez que esta información era revisada, el archivo de datos se enviaba por correo electrónico a la oficina central, donde se consolidaba la información de todas las oficinas regionales.

Encuesta de Hogares MECOVI gestión 2002

Para la Encuesta de Hogares MECOVI 2002, se regresa al sistema de transcripción centralizada en oficina central. Este cambio se debe a que el año pasado, hubo reestructuración de personal en las oficinas regionales.

Al igual que en las gestiones anteriores, la información transcrita se agrupa por Unidades Primarias de Muestreo (UPM's).

En todas las gestiones, los cuestionarios de encuesta contemplan variables precodificadas y variables abiertas (literales), en ambos casos se transcribe esa información tal y como viene de campo. Para el caso de las variables abiertas, se realiza un proceso de codificación posterior para las variables de Ocupación, Actividad, Migración, Productos agrícolas y establecimientos educativos.

IV. Procedimientos para Evaluar la Consistencia de la Información

La consistencia de la información debe ir acompañada inicialmente de una correcta transcripción de datos. Por ello es que se procede a realizar una doble transcripción de los cuestionarios de encuesta, garantizando de este modo una correcta transcripción de datos.

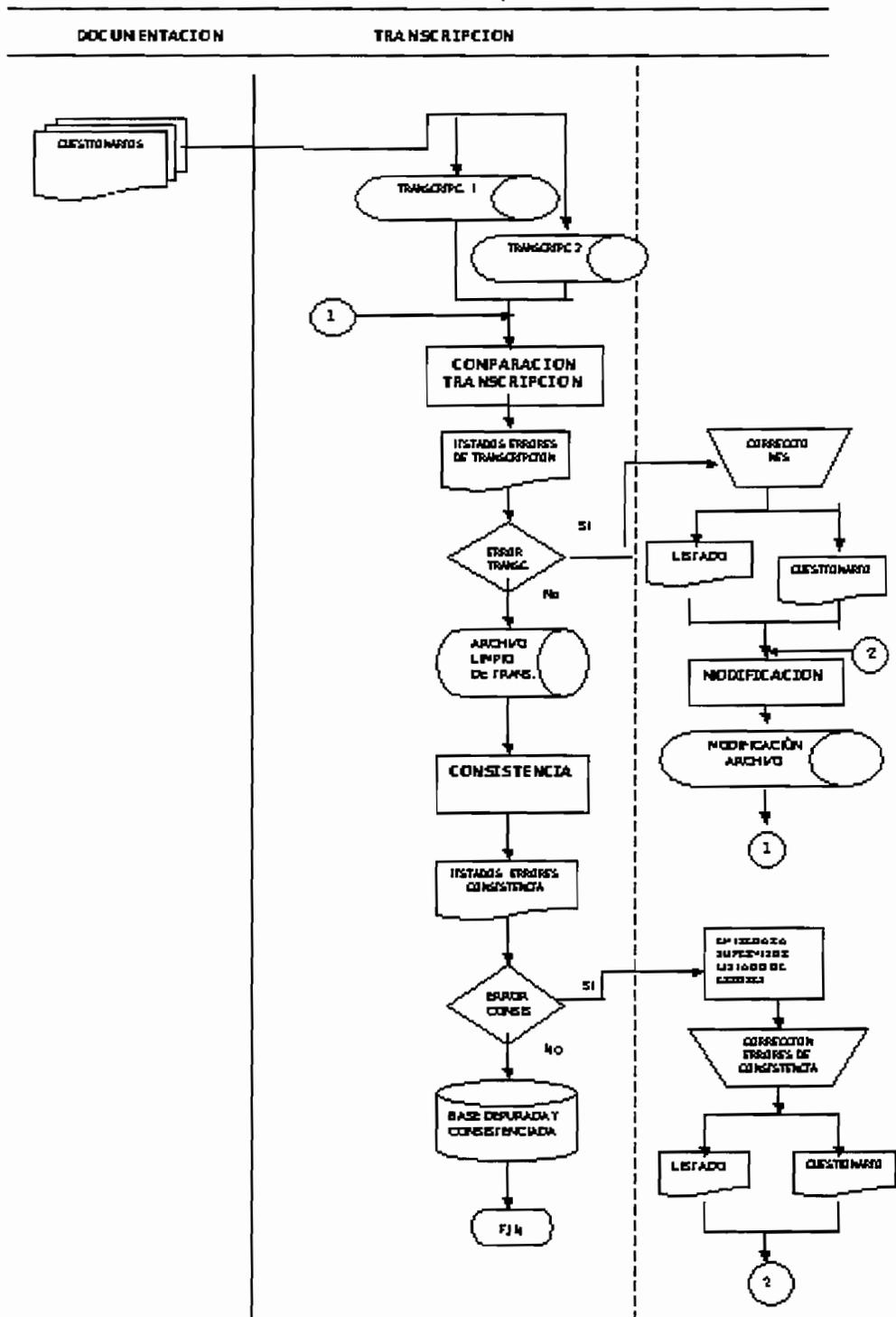
Después de realizar la doble transcripción, se procede a comparar la primera y segunda transcripción, en caso de existir diferencias se generan listados de diferencias, las cuales son corregidas con el cuestionario de encuesta.

Después de corregir las diferencias encontradas, se ejecuta un proceso de consistencia de la información transcrita, emitiendo listados de inconsistencias en caso de encontrarse datos incoherentes.

Los listados de inconsistencias son revisadas por el equipo de Trabajos de Campo, quienes con la ayuda del Supervisor de Campo y encuestadores, analizan las incoherencias y si es preciso se retorna a las vivienda donde se detectó el problema.

La ejecución del proceso de inconsistencias, se constituye en una primera fase de control automático, porque una vez que la información se consolida y se codifica las variables abiertas de Ocupación, Actividad, Migración, Productos agrícolas y establecimientos educativos, se tiene una segunda fase de consistencia de información. En esta segunda fase se controla que las nuevas variables codificadas puedan controlarse con el resto de la información.

DIAGRAMA DE FLUJO DEL PROCESAMIENTO AUTOMÁTICO DE DATOS
ENCUESTA DE HOGARES – MECOVI/2002



En lo que se refiere a la etapa de codificación, para las gestiones 1999 y 2000, se codificó de manera manual, contratando personal de codificación para esta labor. En la gestión 2001, se desarrolló programas de codificación automática, logrando codificar en promedio un 60% de las variables abiertas y el porcentaje restante se codificó de manera manual, directamente en computador. Para los cuestionarios de la gestión 2002, se utilizó una codificación asistida por computador, similar a la utilizada en el Censo de Población y Vivienda 2001.

V. Diseños Computacionales de las Bases de Datos

Hasta la gestión 2001, el software utilizado para la captura de datos y parte de la validación de las Encuestas de Hogares fue I.S.S.A., este software era utilizado debido a que se ajustaba de buena manera en el procesamiento de encuestas de este tipo.

A partir de la gestión 2002, se utilizó el software CPro, paquete que aprovecha las bondades de I.S.S.A. e IMPS. Además de permitir el desarrollo de aplicaciones de Captura de Datos, en tiempos muy cortos en comparación con otros lenguajes de programación. En este software se desarrolló aplicaciones de captura de datos y de validación-consistencia de información.

Para la codificación asistida realizada en la gestión 2002, se desarrolló una aplicación en Visual Basic, utilizando SQL Server y ACCESS. Para ello se exportó la información a ser codificada a estos ambientes. Después de codificarse las variables codificadas fueron pegadas a la Base de Datos original para su utilización completa.

Para la explotación de la información, se exporta los datos a formato SPSS y STATA, para permitir que usuarios internos validen la información antes de poner a disposición de los usuarios externos.

VI. INTEGRACION DE VARIABLES

Una vez que se cuenta con toda la información depurada, se generan variables agregadas de ingreso, gasto y otras. Esta información se obtiene a partir de los módulos de ingreso laboral y no laboral que se aplican a las personas de un hogar, por otro lado el gasto se obtiene a partir de los gastos declarados por cada una de las personas y por el hogar en el caso del consumo alimentos dentro del hogar.

VII. Arquitectura de las Bases de Datos

Una vez que la información es exportada a formato SPSS, La estructura de los archivos de la base de datos es la siguiente:

Nro.	Nombre de Archivo	Descripción
1	HOGAR.sav	Datos a nivel de Hogar
2	POBLACIÓN.sav	Datos a nivel de persona
3	Producción Agrícola.sav	Producción Agrícola
4	Producción Pecuaria.sav	Producción Pecuaria
5	Productos subproductos.sav	Productos derivados y subproductos
6	Consumo de Alimentos.sav	Gastos alimentación dentro del hogar
7	Inventario Bienes Duraderos.sav	Inventario de bienes duraderos

VIII. Generación de Tabulados

Para la generación de tabulados se utilizan principalmente SPSS y STATA.

También se usa CPro, en la revisión de la información en una etapa inicial.

IX. Políticas Institucionales de Acceso a las Bases de Datos

Las Bases de Datos de la Encuesta de Hogares MECOVI 1999, 2000 y 2001 y la documentación respectiva, se encuentran disponibles en el Sitio WEB del INE www.ine.gov.bo , los usuarios pueden acceder a esta información firmando un compromiso donde aceptan hacer un buen uso de la información.

**SISTEMA DE INFORMACIÓN MULTIDIMENSIONAL.
ENCUESTA NACIONAL DE EMPLEO**

INSTITUTO NACIONAL DE ESTADÍSTICA (INE)

CHILE

ÍNDICE

	<u>Página</u>
1. Pantallas de Operación y Descripción de Funcionalidades	61
1.1 Pantalla principal.....	61
1.1.1 Funcionalidad	61
1.2 Selección de consultas multidimensionales	62
1.2.1 Funcionalidad.....	62
1.3 Tabulados frecuentes.....	63
1.3.1 Funcionalidad.....	64
1.4 Consulta predeterminada.....	64
1.4.1 Funcionalidad.....	64
1.5 Creación de slice	64
1.5.1 Funcionalidad.....	65
1.6 Slice generados.....	65
1.6.1 Funcionalidad.....	66
1.7 Modificar slice	66
1.7.1 Funcionalidad.....	67
1.8 Creación de filtro.....	67
1.8.1 Funcionalidad.....	68
1.9 Filtros generados	68
1.9.1 Funcionalidad.....	69
1.10 Modificar filtro.....	69
1.10.1 Funcionalidad.....	70
1.11 Consulta multidimensional.....	70
1.11.1 Funcionalidad.....	71
2. Control de Errores y Descripción	71
2.1 Validación de cuenta y password.....	71
2.2 Seleccionar una dimensión.....	72
2.3 Subir un nivel	72
2.4 Validación de columnas	72
2.5 Validación de filas	72
2.6 Validación de medidas	73
2.7 Validación al generar slice.....	73
2.8 Validación al seleccionar un miembro de un slice.....	73
2.9 Eliminar un slice	74
2.10 Validación en slice y filas	74
2.11 Creación de filtros.....	74
2.12 Generación de tablas	74

Pantallas de Operación y Descripción de Funcionalidades.

1.1 Pantalla principal

La página de inicio permite el ingreso al generador de consultas, el acceso puede ser público en general o también usuarios registrados.

Un perfil es el público, puede ingresar cualquier usuario y realizar consultas directas a la base de datos multidimensional.

Los usuarios registrados deben autenticarse con un login y contraseña, que serán proporcionados por el INE.

The screenshot shows a Microsoft Internet Explorer browser window titled "ENE WEB - Microsoft Internet Explorer". The address bar displays "http://espino.ine.cl/eneweb/paginas/". The main content area contains the following elements:

- Header:** "Encuesta Nacional de Empleo Sistema de Información Multidimensional"
- Logo:** The INE logo, featuring the word "CHILE" in an arc above "INE" and a grid of squares to the right.
- Text:** "Bienvenido a la Base de Datos multidimensional del empleo en Chile. Hemos puesto a su disposición las cifras de la Encuesta Nacional de Empleo desde el trimestre móvil Enero - Marzo de 1996, hasta el último publicado."
- Buttons:**
 - "Ingresar" button under "Todos los Usuarios"
 - "Manual del Sistema" and "Para Registrarse" links
 - "Ingresar" button under "Usuarios Registrados" (next to a login form)
- Login Form:** A table with two rows: "Usuario" and "Contraseña", each with an adjacent input field.

1.1.1 Funcionalidad

Botón Ingresar Todos los Usuarios: Este botón permite el ingreso a la pantalla de selección de consultas multidimensionales con el perfil público. A través de este perfil, cualquier usuario puede ingresar a obtener información de la Encuesta Nacional de Empleo sólo con las dimensiones que le están permitidas.

Botón Ingresar Usuarios Registrados: Este botón realiza la validación del ingreso de usuario y contraseña y permite autenticar al usuario y mostrarle las restricciones de acuerdo al perfil determinado. Este perfil necesita de un usuario y una contraseña que será proporcionada por el INE.

Link Manual de Sistema: A través de ésta pantalla también se puede acceder a un manual de sistema que está definido en formato Word para ayudar al usuario con la navegación para realiza las consultas multidimensionales en el cubo de la Encuesta Nacional de Empleo.

Link Para Registrarse: Existe de igual forma un link para registrarse, se enviará un mail a la dirección ine@ine.cl, en el que se debe solicitar ser usuario del sistema y poder acceder a una contraseña y poder realizar consultas más detalladas.

1.2 Selección de consultas multidimensionales

A través de esta pantalla se puede configurar las consultas a la base de datos multidimensional. Se diseñó la pantalla de esta forma, ya que se quiere representar el área de columnas, área de filas y el área de medidas, como se muestran en Excel. Ej. El usuario deberá seleccionar una dimensión sobre la columna, haciendo doble clic sobre ésta o haciendo clic en la flecha, de igual forma se realiza la operación para filas y medidas.

Se filtran las dimensiones en filas y columnas a medida que se vayan seleccionando y se obtienen las diferentes restricciones que se aplican a cada perfil definido.

Se pueden realizar Slice que son subconjuntos de una Dimensión como por ej. Ocupados o el último trimestre móvil, para obtener una consulta más específica. De igual forma se pueden realizar filtros que son consultas aún mas detalladas ej. Que la población muestral sea mayor que 10.

Encuesta Nacional de Empleo

Si Usted desea crear un **SLICE**
(Subconjunto de una Dimensión)
Si Usted desea crear un **FILTRO**
(Subconjunto de una Dimensión)

Tabulados Frecuentes
(Puede ganar tiempo. Formato HTML.
No requiere acceso a la Base de Datos)

Area de Columnas (Dimensiones)

- ANOS DE ESTUDIO
- AREA ESTIMACION
- AREA GEOGRAFICA
- CATEGORIA OCUPACION
- EDAD
- EST CONYUGAL
- EST OCUPACIONAL
- ESTUDIANTES
- GRUPO OCUPACIONAL
- HRS EFECTIVAS TRABAJO
- HRS NORMALES TRABAJO

Area de Filas (Dimensiones)

- ANOS DE ESTUDIO
- AREA ESTIMACION
- AREA GEOGRAFICA
- CATEGORIA OCUPACION
- EDAD
- EST CONYUGAL
- EST OCUPACIONAL
- ESTUDIANTES
- GRUPO OCUPACIONAL
- HRS EFECTIVAS TRABAJO
- HRS NORMALES TRABAJO

Area de Datos (Medidas)

- POBLACION MUESTRAL
- HRS NORMALES ESTIMADA
- HRS EFECTIVAS ESTIMADA
- SEMANAS BUSCA TRABAJO
- SEMANAS TERMINO TRABAJO
- ANOS ESTUDIOS ESTIMADA
- POBLACION ESTIMADA
- HRS NORMALES MUESTRAL
- HRS EFECTIVAS MUESTRAL
- SEMANAS BUSCA TRABAJO
- SEMANAS TERMINO TRABAJO

Manual del Sistema
Ayuda Metodológica

Generar Tabla Limpia Terminar

Agregar Totales Si No

Independiente del perfil con el que el usuario se haya autenticado, se ingresará a esta pantalla que permite varias funcionalidades. Incluye varios vínculos que se detallan a continuación:

1.2.1 Funcionalidad

Link Manual de Sistema: Al igual que en la pantalla principal se ofrece al usuario un documento Word que le permite ayuda mas detallada de la navegación del sistema.

Link Ayuda Metodológica: Se ofrece al usuario un documento Word que le permitirá, de acuerdo al perfil que tenga, poder realizar consultas multidimensionales y el significado de cada una de las dimensiones, medidas y las restricciones entre filas y columnas.

Link Tabulados Frecuentes: Se ofrece de igual forma, algunos tabulados que están predefinidos y no necesitan acceso a la base de datos multidimensional, estos tabulados son las consultas de la Encuesta Nacional de Empleo más frecuentes que se realizan.

Link Slice: A través de esta pantalla, el usuario puede acotar las consultas a la base de datos multidimensional, se selecciona un subconjunto de alguna dimensión como por ej. Cesantes en el último trimestre móvil.

Link Filtro: Esta pantalla permite obtener información mucho más detallada que los slice. Como por ej. Que las semanas de término de trabajo estimadas sea igual a 5.

Botón Generar Tabla: Este botón realiza diferentes validaciones que permiten la generación una consulta MDX y enviarla a una componente para procesar y desplegar una tabla con la información requerida.

Botón Limpiar: Este botón limpia la pantalla y permite reiniciar y procesar una nueva consulta.

Botón Terminar: Este botón permite terminar la sesion.

Radio Button Agregar Totales: Mediante este radio button, se agregan totales a la consulta generada, se tiene la opción de desplegar los totales, por defecto no se agregan.

1.3 Tabulados frecuentes

Esta pantalla muestra los tabulados más frecuentes de la Encuesta Nacional de Empleo.

Estas consultas no tienen acceso a la base de datos multidimensional por lo que las hace más rápidas de desplegar, ya que están predefinidas en páginas ASP y son estáticas, es decir no se pueden realizar desagregaciones en los miembros de la dimensión.

Tabulados - Microsoft Internet Explorer

Archivo Edición Ver Favoritos Herramientas Ayuda

Tabulados de la Encuesta Nacional de Empleo

- 1a.- POBLACION EN LA FUERZA DE TRABAJO, NIVEL NACIONAL, SEGUN SEXO Y PERIODO (MILES DE PERSONAS)
- 1b.- POBLACION DESOCUPADA, NIVEL NACIONAL, SEGUN SEXO Y PERIODO (MILES DE PERSONAS)
- 2a.- POBLACION EN LA FUERZA DE TRABAJO, SEGUN REGION (MILES DE PERSONAS)
- 2b.- POBLACION DESOCUPADA, SEGUN REGION (MILES DE PERSONAS)
- 3.- OCUPADOS POR RAMA DE ACTIVIDAD ECONOMICA, NIVEL NACIONAL, SEGUN PERIODO (MILES DE PERSONAS)
- 4.- OCUPADOS POR RAMA DE ACTIVIDAD ECONOMICA, SEGUN REGION Y ULTIMO PERIODO (MILES DE PERSONAS)
- 5.- OCUPADOS POR GRUPOS DE OCUPACION, NIVEL NACIONAL, SEGUN PERIODO (MILES DE PERSONAS)
- 6.- OCUPADOS POR GRUPOS DE OCUPACION, SEGUN REGION Y ULTIMO PERIODO (MILES DE PERSONAS)
- 7.- OCUPADOS POR CATEGORIA EN LA OCUPACION, NIVEL NACIONAL, SEGUN PERIODO (MILES DE PERSONAS)
- 8.- OCUPADOS POR CATEGORIA EN LA OCUPACION, SEGUN REGION Y ULTIMO PERIODO (MILES DE PERSONAS)
- 9a.- OCUPADOS POR REGION, SEGUN PERIODO (MILES DE PERSONAS)
- 9b.- DESOCUPADOS POR REGION, SEGUN PERIODO (MILES DE PERSONAS)
- 10.- DESOCUPADOS POR RAMA DE ACTIVIDAD ECONOMICA, NIVEL NACIONAL, SEGUN PERIODO (MILES DE PERSONAS)
- 11.- TASA DE DESOCUPACION, NIVEL NACIONAL Y REGIONAL, SEGUN PERIODO
- 12.- TASA DE DESOCUPACION A NIVEL NACIONAL POR GRUPOS DE EDAD, SEGUN PERIODO

Nota:

- 1.- Período es equivalente a TRIMESTRE
- 2.- Los tabulados frecuentes se presentan en miles de personas y los tabulados generados por el cubo se presentan en valores absolutos.

1.3.1 Funcionalidad

Link Tabulados: Permite obtener una consulta predefinida

Botón Volver: Este botón permite regresar a la pantalla inmediatamente anterior

Botón Terminar: Este botón permite terminar la session

1.4 Consulta predeterminada

Esta es una de las pantallas de las consultas de la Encuesta Nacional de Empleo que se realizan con mayor frecuencia, como se mencionó anteriormente, éstas consultas son estáticas es decir no se puede agregar ni desagregar columnas o filas.

http://espino/eneweb/paginas/TABULADO2.ASP?nombre2=POBLACION TOTAL Y DE 15 AÑOS Y MAS POR SITUACION EN LA FUERZA DE TRABAJO, SEGUN REGION (MILES DE PERSONAS)

Archivo Edición Ver Favoritos Herramientas Ayuda

Atrás Búsqueda Favoritos Historial Ir a

Dirección 20POR%20SITUACION%20EN%20LA%20FUERZA%20DE%20TRABAJO,%20SEGUN%20REGION%20(MILES%20DE%20PERSONAS) Ir a

POBLACION TOTAL Y DE 15 AÑOS Y MAS POR SITUACION EN LA FUERZA DE TRABAJO, SEGUN REGION (MILES DE PERSONAS)

		15 AÑOS Y MAS	MENORES DE 15 AÑOS
HOMBRE	I REGIÓN DE TARAPACÁ	152.329	57.369
	II REGIÓN DE ANTOFAGASTA	175.786	68.309
	III REGIÓN DE ATACAMA	104.052	41.279
	IV REGIÓN DE COQUIMBO	208.176	87.169
	V REGIÓN DE VALPARAÍSO	568.363	215.820
	VI REGIÓN DEL LIBERTADOR GRAL. BERNARDO O´HIGGINS	298.288	112.016
	VII REGIÓN DEL MAULE	336.695	132.322
	VIII REGIÓN DEL BÍO BÍO	699.779	285.386
	IX REGIÓN DE LA ARAUCANÍA	316.930	131.161
	X REGIÓN DE LOS LAGOS	391.780	155.730
	XI REGIÓN DE AISEN, GRAL CARLOS IBÁÑEZ DEL CAMPO	36.205	14.784
	XII REGIÓN DE MACALLANES Y DE LA ANTÁRTICA CHILENA	64.229	20.255
	REGIÓN METROPOLITANA DE SANTIAGO	2.177.461	874.324
MUJER	I REGIÓN DE TARAPACÁ	146.558	55.297
	II REGIÓN DE ANTOFAGASTA	171.147	65.463
	III REGIÓN DE ATACAMA	97.869	39.444
	IV REGIÓN DE COQUIMBO	214.782	84.307
	V REGIÓN DE VALPARAÍSO	604.461	208.220
	VI REGIÓN DEL LIBERTADOR GRAL. BERNARDO O´HIGGINS	290.987	107.930
	VII REGIÓN DEL MAULE	333.703	127.904

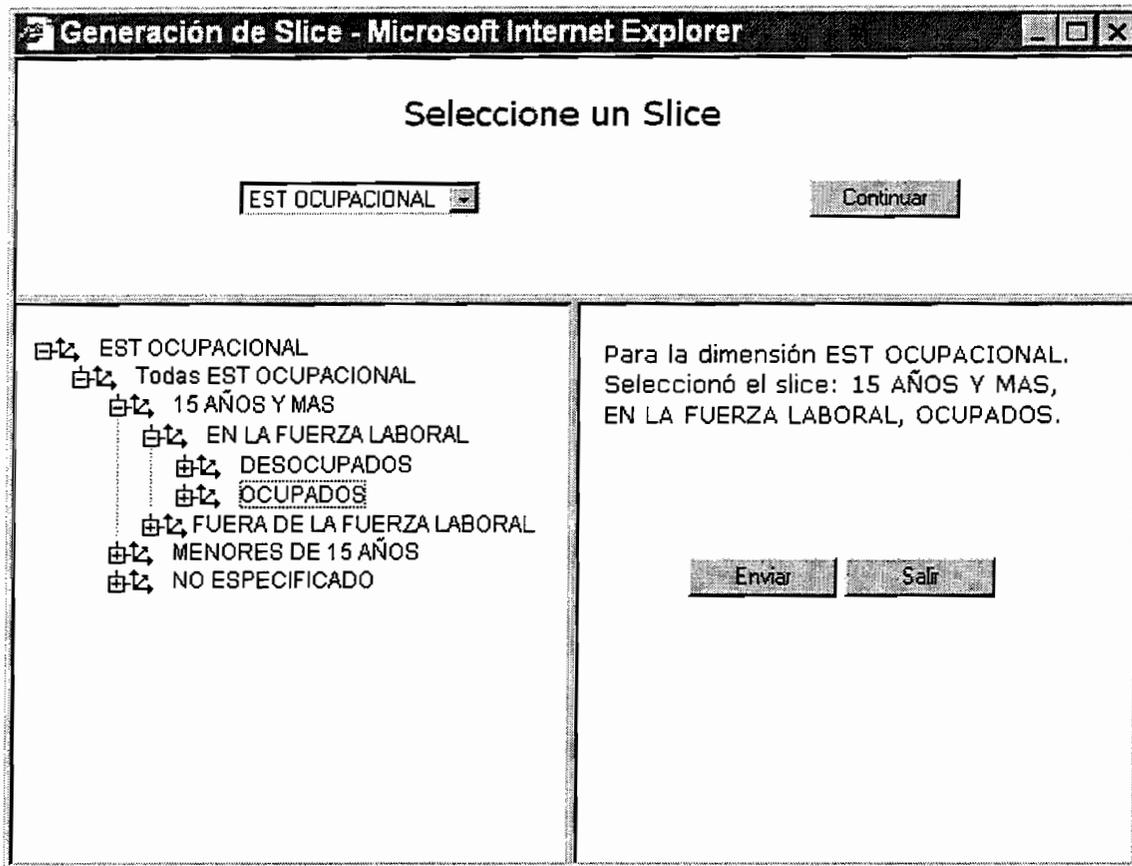
Listo Intranet local

1.4.1 Funcionalidad

Botón Volver: Este botón permite regresar a la pantalla inmediatamente anterior

1.5 Creación de slice

Mediante esta pantalla se puede acotar la consulta a la base de datos multidimensional, es decir, se puede obtener un subconjunto de una dimensión, en este caso tenemos que sólo realizaremos la consulta para los ocupados.



1.5.1 Funcionalidad

Botón Continuar: Este botón permite obtener el archivo XML en el que se muestran todos los miembros de la dimensión elegida.

Link Miembro: Este link permite seleccionar un miembro en particular para generar el slice.

Botón Enviar: Este botón envía el slice generado a la consulta que se generará.

Botón Salir: Este botón permite salir de la pantalla sin generar algún slice.

1.6 Slice generados

Esta pantalla despliega los slice que fueron generados.

Se tiene la opción de modificar el slice haciendo clic en el link del slice o también de eliminarlo chequeando la caja del slice y luego presionando el botón eliminar.

Para terminar se hace click en el botón aceptar.



1.6.1 Funcionalidad

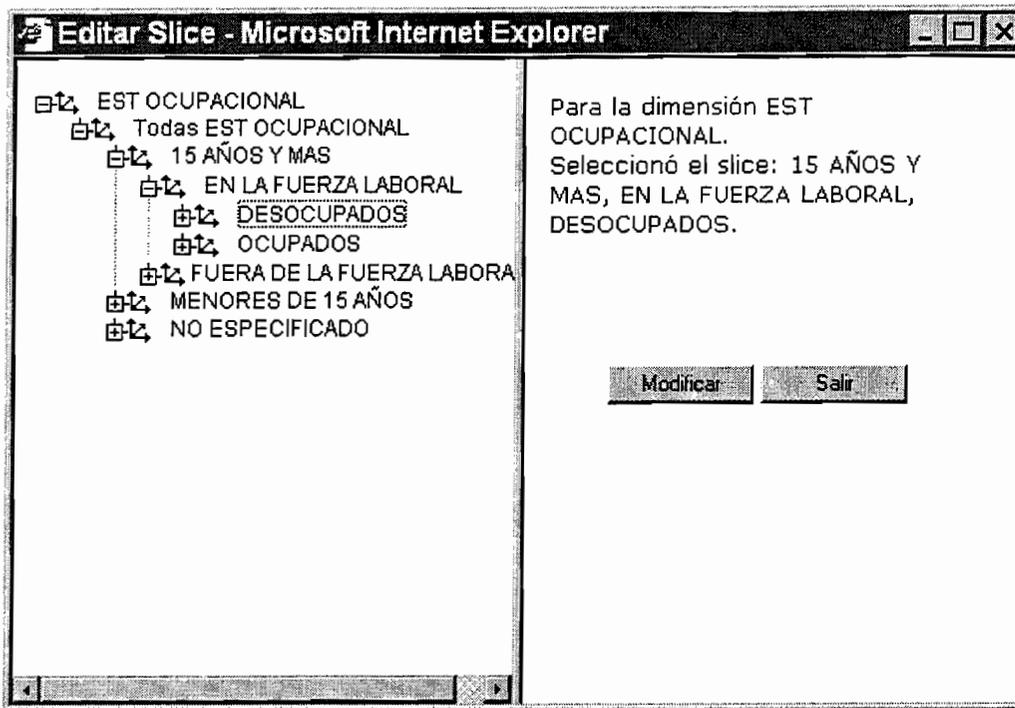
Botón Aceptar: Este botón permite regresar a la pantalla de generación de consultas y enviar los datos modificados

Link Slice: Este link permite modificar el slice que fue generado.

Botón Eliminar: Este botón elimina un slice, previamente se debe seleccionar chequeando en la caja del slice.

1.7 Modificar slice

Esta pantalla permite modificar un slice generado, en esta pantalla sólo se muestra la dimensión que sobre la que se generó el slice, al modificar el slice, se realiza la misma operación que cuando se genera un slice por primera vez.



1.7.1 Funcionalidad

Link Miembro: Este link permite seleccionar un miembro en particular para modificar el slice.

Botón Enviar: Este botón envía el slice modificado a la consulta que se generará.

Botón Salir: Este botón permite salir de la pantalla sin modificar algún slice.

1.8 Creación de filtro

Mediante esta opción se puede obtener una consulta aún mas detallada de alguna dimensión en particular, en este caso tenemos que se realizará un filtro para que la población estimada sea igual a cinco.

FILTROS - Microsoft Internet Explorer

Seleccione un Filtro

Medidas
Poblacion Estimada

Operadores Lógicos
Igual (=)

Valor
5

Continuar

Seleccionó el siguiente filtro:
Para Poblacion Estimada=5

Enviar Salir

1.8.1 Funcionalidad

Botón Continuar: Este botón permite generar el filtro con la medida el operador lógico y el valor.

Botón Enviar: Este botón envía el filtro generado a la consulta que se generará.

Botón Salir: Este botón permite salir de la pantalla sin generar algún filtro.

1.9 Filtros generados

Esta pantalla despliega los filtros que fueron generados.

Se tiene la opción de modificar el filtro haciendo clic en el link del filtro o también de eliminarlo chequeando la caja del filtro y luego presionando el botón eliminar.

Para terminar se hace click en el botón aceptar.

FILTROS - Microsoft Internet Explorer

Seleccione un Filtro

Medidas
Poblacion Estimada

Operadores Lógicos
Mayor (>)

Valor
5

Continuar

Seleccionó el siguiente filtro:
Para Poblacion Estimada>5

Enviar Salir

1.9.1 Funcionalidad

Botón Aceptar: Este botón permite regresar a la pantalla de generación de consultas y enviar los datos modificados

Link Filtro: Este link permite modificar el filtro que fue generado.

Botón Eliminar: Este botón elimina un filtro, previamente se debe seleccionar chequeando en la caja del filtro.

1.10 Modificar filtro

Esta pantalla permite modificar un filtro generado, al modificar el filtro, se realiza la misma operación que cuando se genera un filtro por primera vez.

FILTROS - Microsoft Internet Explorer

Modificar Filtro

Medidas
Poblacion Estimada

Operadores Lógicos
Mayor o Igual (>=)

Valor
10

Continuar

Seleccionó el siguiente filtro:
Para Poblacion Estimada >= 10

Enviar Salir

1.10.1 Funcionalidad

Botón Continuar: Este botón permite modificar el filtro con la medida el operador lógico y el valor.

Botón Enviar: Este botón envía el filtro modificado a la consulta que se generará.

Botón Salir: Este botón permite salir de la pantalla sin modificar algún filtro.

1.11 Consulta multidimensional

Esta es la pantalla de respuesta a una consulta a la base de datos multidimensional, se despliega las columnas, filas y medidas que fueron seleccionadas, de igual forma se muestra los slice y los filtros que se generaron para consulta final.

Esta tabla tiene la particularidad de ser dinámica, es decir se puede agregar o desagregar una fila o columna si es que poseen miembros internos. Al pasar el mouse sobre las filas o columnas se mostrará el nombre subrayado, es decir se debe hacer clic en el link y se desagregará la fila o columna seleccionada.

Se tiene la opción de exportar a Excel.

http://espino/eneweb/paginas/tabla_ene_web.asp?filas=1&columnas=1&str_Filas=[PARENTESCO] Todas Microsoft Internet Explorer

Archivo Edición Ver Favoritos Herramientas Ayuda

ENCUESTA NACIONAL DE EMPLEO

Area de Columnas : SEXO
 Area de Filas : PARENTESCO
 Area de Medidas : POBLACION ESTIMADA

	HOMBRE	MUJER	Total SEXO
CONVIVIENTE	3.421.625	20.485.432	23.907.057
CONYUGE	4.016.490	178.799.700	182.816.190
HERMAND(A) / CUÑADO(A)	7.930.894	8.547.680	16.478.574
HIJO(A) HIJASTRO(A)	251.185.725	233.249.534	484.435.259
JEFE DE HOGAR	239.594.526	69.269.922	308.864.448
NIETO(A)	44.566.522	42.577.160	87.143.682
NO PARIENTE	3.977.492	3.314.181	7.291.673
OTRO PARIENTE	10.505.459	10.349.268	20.854.727
PADRES / SUEGROS	2.564.795	10.727.480	13.292.275
SERVICIO DOMESTICO	131.387	4.716.342	4.847.729
YERNO/NUERA	10.640.829	7.983.738	18.624.567
Total PARENTESCO	578.535.744	590.020.437	1.168.556.181

Si desea exportar a planilla Excel presione [Aquí](#)

1.11.1 Funcionalidad

Link Aquí: Este link permite exportar a excel la consulta generada.

Botón Volver: Este botón permite volver a la pantalla inmediatamente anterior.

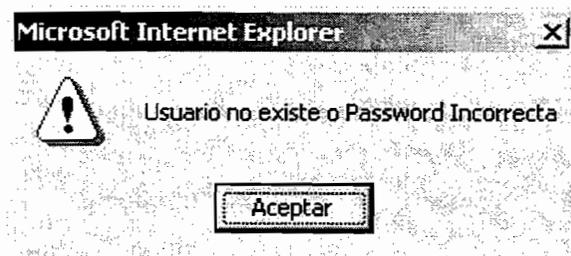
Botón Nueva Consulta: Este botón vuelve a la pantalla de selección de consultas multidimensionales.

Botón Terminar: Este botón termina la session.

2 Control de Errores y Descripción.

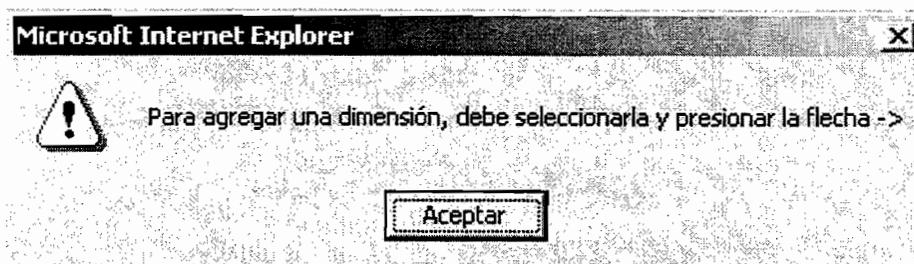
2.1 Validación de cuenta y password

Al ingresar una cuenta de usuario o una password incorrecta se despliega el siguiente error:



2.2 Seleccionar una dimensión

Si en cualquier área se hace clic en la flecha de la derecha o de la izquierda y previamente no se seleccionó alguna dimensión, se despliega el siguiente mensaje:



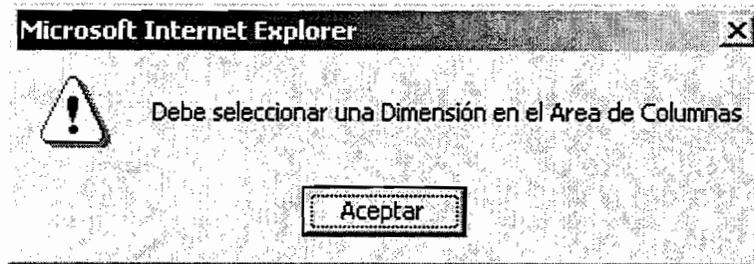
2.3 Subir un nivel

Si en cualquier área se hace clic en la flecha de arriba o abajo y sólo existe una dimensión, se despliega el siguiente mensaje:



2.4 Validación de columnas

Si al generar la tabla no se seleccionó alguna dimensión en el área de columnas, se despliega el siguiente error:



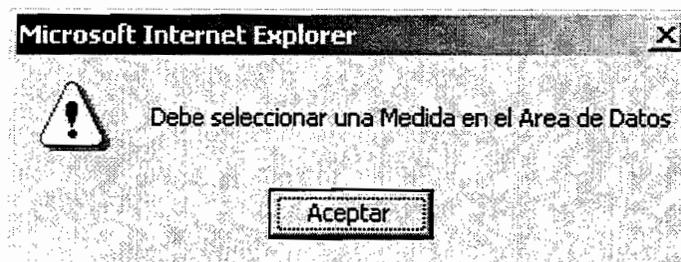
2.5 Validación de filas

Si al generar la tabla no se seleccionó alguna dimensión en el área de filas, se despliega el siguiente error:



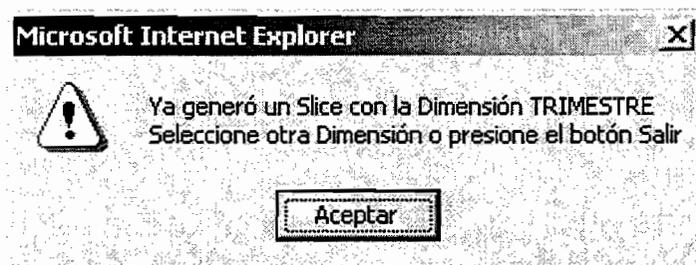
2.6 Validación de medidas

Si al generar la tabla no se seleccionó alguna dimensión en el área de filas, se despliega el siguiente error:



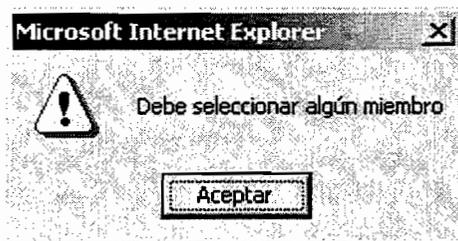
2.7 Validación al generar slice

Si se desea generar un slice y ya fué creado un slice previo con la misma dimensión, entonces se despliega el siguiente mensaje:



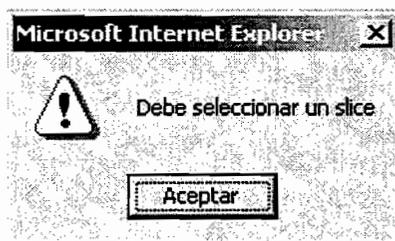
2.8 Validación al seleccionar un miembro de un slice

Si se desea generar un slice y ya y no se ha seleccionado ningún miembro y se presiona el botón enviar, se despliega el siguiente mensaje:



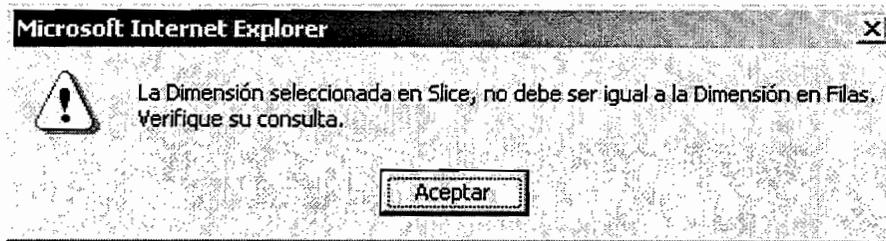
2.9 Eliminar un slice

Si se desea eliminar un slice y no se ha seleccionado ninguno y se presiona el botón eliminar, se despliega el siguiente mensaje:



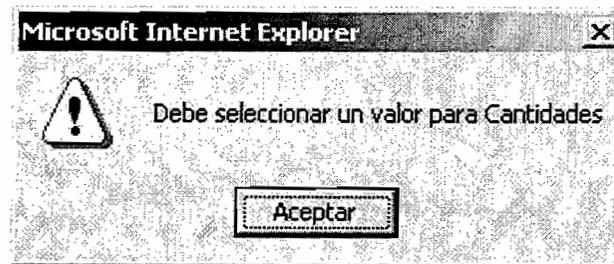
2.10 Validación en slice y filas

Si al generar la tabla se creó un slice y en el área de filas o columnas también existe la misma dimensión, se despliega el siguiente mensaje:



2.11 Creación de filtros

Si al generar un filtro se presiona en botón Continuar y no se ha ingresado un valor para cantidades, se despliega el siguiente mensaje:



2.12 Generación de tablas

Si al generar la tabla y no se obtienen datos a la consulta seleccionada, entonces se despliega el siguiente mensaje:

http://espino/eneweb/paginas/tabla_ene_web.asp?filas=1&columnas=1&str_filas=[CATEGORIA OCUPACIONAL] ...

Archivo Edición Ver Favoritos Herramientas Ayuda

Atrás Búsqueda Favoritos Historial

Dirección STRE],[Todas%20TRIMESTRE],[2002%20%20%20JUN-AGO]);&txtMember=([Measures],[Poblacion%20Estimada])%20=%201; Ir a

ENCUESTA NACIONAL DE EMPLEO

Area de Columnas : EST OCUPACIONAL
Area de Filas : CATEGORIA OCUPACIONAL
Area de Medidas : POBLACION ESTIMADA
Slice : Para la dimensión TRIMESTRE.
Selecciónó el slice: 2002 JUN-AGO
Filtros : (Poblacion Estimada) = 1

No se obtuvieron datos para esta consulta.

Volver Nueva Consulta Terminar

Listo Intranet local

**INFORME METODOLÓGICO INFORMÁTICO
ENCUESTA DE HOGARES, AÑO 2003**

DEPARTAMENTO ADMINISTRATIVO NACIONAL DE ESTADÍSTICAS (DANE)

COLOMBIA

ÍNDICE

	<u>Página</u>
Introducción	79
Aspectos Generales	79
Procesamiento de Datos	80
Anexo 1	82
Modelo E-R del Sistema de Captura de la ECH	82
Anexo 2	83
Diccionario de Datos.....	83

Introducción

En Colombia, el DANE tiene la responsabilidad de la planeación, levantamiento, procesamiento, análisis y difusión de las estadísticas oficiales que se distribuyen, a nivel nacional y para los organismos internacionales de los cuales el país hace parte como: la Organización de las Naciones Unidas y sus respectivas agencias, el Fondo Monetario Internacional (FMI), la Organización de los Estados Americanos (OEA), la Organización Mundial del Comercio (OMC), la Oficina Internacional del Trabajo (OIT), entre otras

Para cumplir con esta responsabilidad en lo que compete a las estadísticas de carácter social, el DANE cuenta con encuestas dirigidas a los hogares como son: la Encuesta Continua de Hogares, que mide el mercado laboral colombiano; la Encuesta Nacional de Calidad de Vida, que mide las condiciones de vida de los colombianos; la Encuesta Nacional de Ingresos y Gastos; la encuesta de Trabajo Infantil; etc.

Estos proyectos se trabajan de forma independiente y cada uno de ellos cuenta con sistemas de información dentro de los cuales se encuentran los módulos de captura (descentralizada), procesamiento (centralizado) y sistemas de difusión (impreso y electrónicos).

En este documento se explicarán los sistemas de información con que cuenta la Encuesta Continua de Hogares.

Aspectos Generales

Para el Sistema de la encuesta de Hogares se desarrollaron sistemas independientes así:

- Sistema de captura con validación:
 - Desarrollado en Visual fox-Pro.
 - Monousuario.
 - Cuenta con los módulos de control de la muestra, captura de formularios, inconsistencias, rendimientos, control de cobertura, entre otros.
 - Un digitador por Sede o Subsele, con experiencia en digitación y conocimiento de la encuesta.
 - En el anexo 1 se muestra el modelo E-R de una etapa particular de la encuesta.
 - No se hace codificación automática. Para la codificación de ocupaciones se utiliza la CON y para la codificación de ramas de actividad se utiliza la CIU-3.
- Sistema de procesamiento:
 - Desarrollado en SAS.
 - Cliente/Servidor.
 - Cuenta con los módulos de generación de factor de cobertura, generación de factores de expansión (PT y PET), generación de plano y librerías SAS para los usuarios finales, generación de cuadros de salida y generación de coeficientes de variación.
 - En el anexo 2 se muestra el diccionario de datos entregado a los usuarios finales para el manejo del archivo plano.
 - Los archivos del nivel de microdato (conservando la reserva estadística), son de uso restringido y se comercializa por intermedio del Banco de Datos (Dirección de Mercadeo).
- Sistema de difusión con base de datos agregados:

- Desarrollado en ORACLE.
- Cliente/Servidor y WEB
- Cuenta con los módulos de cargue de información agregada y cargue de fichas metodológicas.
- El uso de la información agregada es libre y no tiene costo alguno.

Procesamiento de Datos

Depuración y corrección de inconsistencias: Una vez consolidada la información que llega de la Sedes y subsedes se revisa la información, utilizando un programa adicional para generación de inconsistencias. El porcentaje de error encontrado es por debajo del 0.5%. Sin embargo, se corrige en su totalidad volviendo en muchas ocasiones a la fuente si es necesario.

Factor de cobertura: Una vez depurado al 100% el archivo de datos inicial, a cada segmento se le listan el total de encuestas esperadas sobre el total de encuesta efectivas y se hace una ajuste por cobertura, el cual se utiliza posteriormente en el factor de expansión.

Pesos de submuestreo: Para cada segmento y de acuerdo con lo reportado por la Sedes y Subsedes de lo encontrado en el recuento, se hace necesario en algunas oportunidades, aplicar un submuestreo al segmento por envejecimiento de la muestra, el cual se utiliza posteriormente para el factor de expansión.

Ajuste del factor de expansión: La Dirección de Censos y Demografía entrega para cada municipio de la encuesta, por Cabecera – Resto, y por rangos de edad las proyecciones de población de cada 15 días. Esto implica, que el ajuste por proyecciones de población para generar el factor de expansión final se de en dos pasos que se describen a continuación:

- Ajuste con respecto a la población total: Se cuenta todas las personas de la muestra a las que se les has aplicado los factores iniciales, los de cobertura y los de pesos de submuestreo por segmento, se suman y se ajustan de acuerdo con los datos de la nueva población.
- Ajuste con respecto a la población en edad de trabajar: Se separan las personas de 10 años y más (zona resto) y de 12 años y más (zona cabecera), de los menores (población que no está en edad de trabajar) y se aplica un nuevo ajuste de acuerdo con las proyecciones dadas para estos rangos de edad.
- Este procedimiento hace que se utilicen dos factores de expansión: uno para los análisis de hogares y otro para los análisis de personas.

Con este procedimiento de ajuste por rangos de edad, se garantiza que los saltos en la población en edad de trabajar no estén sesgados por el ajuste de población total.

Generación de cuadros de salida: con el PROC TABULATE de SAS y el ODS HTML, se generan los cuadros de salida de la encuesta de acuerdo con el período de publicación, la agregación de la muestra y la cobertura. Los cuadros se generan en formato HTML y luego se pasan a EXCEL. Estos se entregan al usuario final por intermedio del banco de datos y son comercializados.

- Cuadros mensuales: 15 cuadros para el total nacional y total 13 áreas metropolitanas.
- Cuadros bimestrales: 15 con un nivel de desagregación mayor.
- Cuadros trimestrales: 120 cuadros que se generan para toda la cobertura de la encuesta
- Cuadros semestrales: 120 cuadros que se generan para toda la cobertura de la encuesta desagregando por región.
- Cuadros anuales: 80 cuadros generados por departamento

Generación de errores de muestreo: Para los cuatro principales indicadores se generan los coeficientes de variación y los respectivos intervalos de confianza con los errores estándar relativo.

Generación de archivos para cargue de la BDA: Una vez se han publicado los datos trimestrales, se generan los archivos planos para cargar a la base de datos agregados que se tiene para el sistema de difusión en ambiente cliente/servidor y en WEB.

ANEXO 2 – DICCIONARIO DE DATOS
SISTEMA DE PROCESAMIENTO Y DIVULGACION DE
LA ENCUESTA CONTINUA DE HOGARES

1. La Encuesta Continua de Hogares produce tres archivos mensuales
 - Total 13 áreas
 - Total Cabecera
 - Total Resto
2. Cada uno de los anteriores archivos tiene un diseño de registro
3. Resultados Mensuales que se pueden sacar según la muestra de la Encuesta
 - Total 13 áreas, sin desagregar ninguna área o ciudad
 - Total Nacional sin desagregar Cabecera o Resto
4. Resultados Bimestrales que se pueden sacar según la muestra de la Encuesta
 - Total 13 áreas, sin desagregar ninguna área o ciudad
 - Total 9 áreas, sin desagregar ninguna área o ciudad
 - Total 4 áreas, desagregando: Medellín área metropolitana (AM), Barranquilla AM, Bogotá y Cali.
 - Total Nacional, Cabecera y Resto

NOTA: para el bimestre se deben unir dos meses y tomar el Factor de Expansión de cada mes y dividirlo por dos
5. Resultados Trimestrales que se pueden sacar según la muestra de la Encuesta:
 - Total 13 áreas y total por cada una de las áreas o ciudades
 - Total Nacional, Cabecera y Resto

NOTA: para el trimestre se deben unir tres meses y tomar el Factor de Expansión de cada mes y dividirlo por tres.
6. Resultados Semestrales que se pueden sacar según la muestra de la Encuesta:
 - Total 13 áreas y total por cada una de las áreas o ciudades
 - Total Nacional por regiones, Cabecera por regiones

NOTA: para el semestre se deben unir seis meses y tomar el Factor de Expansión de cada mes y dividirlo por seis.
7. Resultados Anuales que se pueden sacar según la muestra de la Encuesta:
 - Total 13 áreas y total por cada una de las áreas o ciudades
 - Total Nacional por regiones, Cabecera por regiones y Resto por regiones

NOTA: para el año se deben unir doce meses y tomar el Factor de Expansión de cada mes y dividirlo por doce.

NOTA ACLARATORIA

El DANE no se hace responsable de los datos que se procesen o divulguen fuera de los dominios establecidos.

**ENCUESTA DE HOGARES DE
PROPÓSITOS MÚLTIPLES**

CARLOS LORÍA LORÍA

INSTITUTO NACIONAL DE ESTADÍSTICA Y CENSOS (INEC)

COSTA RICA

ÍNDICE

	<u>Página</u>
I. Encuesta de Hogares de Propósitos Múltiples de Costa Rica	87
Objetivos de la encuesta.....	87
Características de la encuesta.....	87
Diseño muestral.....	87
Tamaño de la muestra	87
Contenido temático	87
II. Estrategia para la Captura de Datos.....	88
Reseña histórica	88
Proceso actual.....	88
Codificación de preguntas.....	89
III. Procedimientos para Evaluar la Consistencia.....	89
IV. Diseños Computacionales de las Bases de Datos.....	90
V. Métodos para la Imputación de Datos	90
VI. Integración de Variables.....	91
VII. Arquitectura de las Bases de Datos	92
VIII. Generación de Tabulados	92
IX. Políticas Institucionales de Acceso a las Bases de Datos.....	92
Anexo 1	94
Plan de Inconsistencias	94
Anexo 2	97
Plan de Creación de Variables	97

I. Encuesta de Hogares de Propósitos Múltiples de Costa Rica

Costa Rica cuenta con una larga trayectoria en la realización de Encuestas de Hogares, esta se inició en 1966 con el programa de la Encuesta Centroamericana de Hogares, este programa permaneció hasta 1971, año en que fue discontinuado por dificultades financieras. En el año 1976 y hasta 1986 se realizó la Encuesta Nacional de Hogares, Empleo y Desempleo, en un esfuerzo conjunto del Ministerio de Trabajo y de Seguridad Social y de la entonces Dirección General de Estadística y Censos.

Entre los años 1985 y 1986 se realiza una revisión y rediseño integral del Programa Permanente de Encuestas de Hogares, dando origen al Programa de Encuestas de Hogares de Propósitos Múltiples actual que se realiza desde 1987 en julio de cada año. Este programa es desarrollado conjuntamente por el Instituto Nacional de Estadística y Censos y cuenta además con el apoyo de diferentes instituciones que solicitan módulos especiales.

La EHPM constituye una valiosa fuente de información sobre diversos aspectos relativos a los hogares y a las personas que conforman esos hogares. Tiene especial importancia el Módulo de Empleo que constituye el módulo principal de la investigación, el cual indaga sobre el empleo, el desempleo, el subempleo y los ingresos, en general sobre aspectos referentes a la fuerza de trabajo y sus características. La encuesta sirve también como medio para investigar otros temas complementarios acerca de las características demográficas y socioeconómicas de los hogares y de sus residentes.

Objetivos de la Encuesta

- a) Mantener un flujo continuo de estadísticas relacionadas con la fuerza de trabajo, el empleo, el desempleo, el subempleo y los ingresos, así como de otras variables socioeconómicas necesarias para el establecimiento de políticas y la formulación de planes orientados al desarrollo económico y social del país, y para la evaluación de sus efectos.
- b) Proveer información periódica, sistemática y oportuna en los períodos ínter censales, referente a las variables mencionadas.
- c) Servir de fuente de información a instituciones gubernamentales, universitarias, o de investigación, interesadas en temas relativos a la población y el empleo, y en otros temas que se introduzcan periódicamente en la encuesta.

Características de la Encuesta: La Encuesta de Hogares de Propósitos Múltiples tiene cobertura nacional y la muestra permite obtener estimaciones confiables a nivel de Región de Planificación.

Diseño Muestral: El diseño muestral corresponde a un diseño probabilístico de áreas, estratificado y bietápico. Considera doce estratos para fines de selección y distribución de la muestra - cada región de planificación divide por zona urbana y rural -; en una primera etapa se seleccionan segmentos censales o Unidades Primarias de Muestreo (UPM), y en una segunda etapa se seleccionan Unidades Secundarias de Muestreo (USM) dentro de las UPM seleccionadas en la primera etapa.

Tamaño de la Muestra: El tamaño de la muestra inicial era de **10890** viviendas el cual ha aumentado a **13175** en el año 2002, por el crecimiento en el número de viviendas.

Contenido Temático: La temática investigada por la Encuesta de Hogares de Propósitos Múltiples está asociada a temas tales como la población total según los diferentes grupos de edad, fuerza de trabajo, empleo, desempleo, subempleo, ingresos, pobreza, inactividad y niveles de escolaridad, además de los módulos especiales que solicitan los usuarios sobre vivienda, servicios de salud, consumo aparente de alimentos, servicios de telecomunicaciones, trabajo infantil y adolescente, adulto mayor, etc. En general las variables se encuentran desagregadas por sexo, por región de planificación y por zona urbana y rural del país.

II. Estrategias para la Captura de Datos

Reseña Histórica: Desde que se inició el actual programa de encuestas en el año 1987, una vez que se termina el trabajo de campo los cartapacios conteniendo los cuestionarios de la encuesta son traídos a las oficinas centrales del INEC, donde son criticados y codificados, posteriormente estos son enviados al Proceso de Digitación donde se da inicio al proceso de transcribir la información. En un principio la captura de los datos se realizaba en estaciones IBM 3742, en donde la información era grabada en un disquete que posteriormente se pasaba a una cinta magnética para ser procesado en un computador IBM 4361.

En el año de 1995 la entonces Dirección General de Estadística y Censos adquirió un grupo de microcomputadoras IBM Aptiva y se decide realizar el proceso de captura de datos en estas máquinas. Para este propósito se utilizó el Integrated Microcomputer Processing System (**IMPS**) desarrollado por el International Statistical Programs Center (ISPC) de la Oficina del Censo de los Estados Unidos.

Proceso actual: El IMPS para la captura de datos se utilizó hasta la encuesta del año 2002, es entonces que se decide cambiar por el Census and Survey Processing System (**CSPRO**) que combina las características del IMPS y del software para análisis estadístico ISSA, en un ambiente Windows.

El programa de captura de datos actual esta formado por una serie de pantallas con un diseño similar al del cuestionario utilizado para la recolección de datos, debido a que el cuestionario esta formado por varios registros divididos por temas, la entrada de datos se hizo siguiendo ese formato. Los registros que forman la captura de datos son: carátula, vivienda y servicios, sociodemográficas y socioeconómicas. El programa de captura esta diseñado para que ejecuten automáticamente todos los saltos que tiene el cuestionario y además se agregaron una serie de saltos manuales para agilizar el proceso a los digitadores.

Además del programa de captura hecho en CSPRO, se utilizan dos programas diseñados en FoxPro para realizar tareas como la asignación de los cartapacios a los digitadores y verificadores y otro para revisar la cobertura de la encuesta.

El programa de **Gestión de Archivos** permite al encargado del proceso de digitación enviar a la máquina de cada digitador información sobre el número de segmento con que le toca trabajar, de esta manera automáticamente al ejecutar la aplicación aparecerá el número de segmento, con esto se disminuye el riesgo de que el operador digite mal la identificación de todo el grupo de boletas.

Una vez que el digitador termine con el segmento asignado, este se lo entrega al Supervisor que a su vez “*cierra*” la operación de captura, el programa de gestión de archivos hace una copia del segmento digitado y lo envía a una carpeta de denominada **digitados** en el servidor de datos.

Para la verificación el proceso es el mismo, el Supervisor envía el segmento que se encuentra digitado en el servidor a la máquina del verificador, una vez que este termina de verificar el segmento, el supervisor lo envía al servidor a una carpeta denominada **verificados**.

Una vez realizado el proceso anterior, se aplica el programa denominado **Control de Cobertura**. El procedimiento es el siguiente: una vez que los segmentos son enviados al servidor a la carpeta de **verificados**, este chequea que estén todos los segmentos seleccionados y que no falten ni sobren viviendas, de acuerdo a una tabla que contiene la distribución de la muestra que proporciona el Proceso de Muestreo. En caso de que encontrara algún problema, el programa envía a una carpeta llamada **pendientes** con el propósito de que sean revisados, por el contrario, sino encuentra errores los segmentos son enviados a una carpeta denominada **definitivos** donde quedan listos para la siguiente etapa que es la de detección y corrección de errores.

Codificación de preguntas: En cuanto a la codificación de las preguntas debemos señalar que el cuestionario de la Encuesta de Hogares de Propósitos Múltiples esta formado por varios tipos de preguntas:

1. **Marcar con equis (X):**

- **Marcar una alternativa:** en este caso se debe escoger una sola alternativa, y la respuesta está precodificada.
- **Marcar más de una alternativa:** se marcan varias opciones con equis (X), también en este caso las respuestas están precodificadas.

2. **Especificar una respuesta:**

- **De forma alfabética:** en algunas preguntas aparece un espacio con líneas continuas donde el entrevistador debe anotar la respuesta que dé el informante, posteriormente en la etapa de codificación se le asignará el respectivo código, labor que se realiza en la Oficina.
- **De forma numérica:** en otro grupo de preguntas aparecen casillas abiertas donde se anotan datos numéricos que generalmente corresponden a montos de dinero y que son digitados directamente.

III. Procedimientos para Evaluar la Consistencia

El procedimiento de detección y corrección de errores para la Encuesta de Hogares inicia cuando se recibe por parte del Area de Censos y Encuestas el “Plan de Inconsistencias para la Encuesta de Hogares”, el cual consiste en un listado con todas las variables del cuestionario y las indicaciones para detectar los errores.

Una vez recibido este documento se procede a programar dentro del modulo **Batch Edit** del **CSPro** todas las reglas de consistencia. Básicamente lo que el programa hace es detectar los errores y mandar mensajes, esto con el fin de que el grupo encargado de la limpieza realice la corrección correspondiente. Los encargados de la limpieza de los datos son los supervisores de la Encuesta.

En el momento de la captura de los datos se validan los rangos de la mayoría de las variables, se excluyen de esta validación las preguntas correspondientes a ingresos ya que ahí se permite ingresar cualquier valor. Posteriormente, en la etapa de detección de errores los rangos permitidos para las diferentes variables de ingreso son chequeados.

El procedimiento de limpieza consiste en que una vez digitados y verificados los segmentos de la encuesta se agrupan en lotes de diez segmentos que son entregados a cada uno de los grupos encargados de la limpieza, estos grupos están formados por dos personas, en donde una de las personas se encarga de buscar el cuestionario que contiene el error y la otra de hacer la corrección en la computadora.

El encargado de hacer la corrección en la computadora tendrá abiertas dos pantallas, en una la captura de datos y en otra la salida que genera el programa de inconsistencias en donde se lista la identificación del caso a revisar (segmento, vivienda, hogar, número de línea, y la variable con error) para facilitar su búsqueda. La corrección se hace sobre el archivo original.

Durante la captura no se forman archivos por temas, todo se digita y se depura simultáneamente dentro del mismo archivo; solo en el caso de que el tema sea muy especializado, se generan a partir del archivo original otros archivos de datos que se entregan a las personas que solicitaron el módulo con el fin de que ellos realicen en conjunto con personal del INEC la limpieza de la información, esto se da por ejemplo con el Módulo de Consumo Aparente de Alimentos, el cual es revisado por nutricionistas, o el Módulo de Trabajo Infante Juvenil, el cual fue revisado en conjunto con personal del Ministerio de Trabajo.

En lo referente a los tiempos necesarios para el proceso se tiene que para la elaboración del Plan de Inconsistencias se utilizan cinco semanas y lo trabajan dos personas; la programación de este plan se realiza en nueve semanas y una persona es la encargada de esta tarea; la ejecución de la limpieza de los datos dura tres semanas y la ejecutan doce personas y un supervisor.

IV. Diseños Computacionales de las Bases de Datos

Para el proceso de captura así como para el proceso de detección de errores se utiliza el sistema integrado **CSPro**, en el se han desarrollado las aplicaciones necesarias para estos efectos. Aparte de usar el **CSPro**, se utilizan algunos programas hechos en FoxPro con el propósito de facilitar la manipulación de los segmentos a la hora de su captura y para revisar la cobertura.

Para los usuarios finales de la información, una vez limpio el archivo y agregadas las variables se pone a disposición de ellos en formato DBF, o en el formato que el usuario lo solicite, principalmente se usa el formato SAV para ser usado en el SPSS.

El INEC no cuenta actualmente con software de desarrollo propio para el procesamiento de datos debido a que no dispone de recursos financieros ni humanos para ello, y a que, hasta ahora, se ha considerado suficiente y con buenos resultados la aplicación del IMPS y el **CSPro**, al menos para las fases de entrada, depuración y tabulación. No obstante hay ausencia de un sistema de generación y consulta más ágil para usuarios internos y externos.

V. Métodos para la Imputación de Datos

La Encuesta de Hogares de Propósitos Múltiples contempla en su diseño muestral un ajuste al factor de expansión en caso de existir No Respuesta Total, este ajuste consiste en “compensar” la pérdida de muestra debido a las viviendas que no respondieron, de esta forma se le da un peso mayor a las viviendas que si respondieron con el fin que estas representen el total de viviendas que debieron ser entrevistadas.

La No Respuesta Parcial se analiza en el Proceso de Crítica y Codificación, en donde se revisan los cuestionarios con información faltante y se determina si esta puede ser completada por vía telefónica, en caso de que no se pueda completar por este medio se le asigna a la información faltante el código de ignorado (9). Si la información faltante es mucha se le cambia el código de resultado de entrevista al cuestionario y este es tratado como No Respuesta Total y se ajusta por el factor de expansión.

La imputación por medio de programa no es una práctica común en la Encuesta de Hogares de Costa Rica. Este procedimiento si se utilizó en el último Censo Nacional de Población y Vivienda del año 2000, donde se utilizó la técnica de imputación dinámica “Hot Deck”. Como se mencionó anteriormente la información faltante no se imputa sino que se trata como pérdida de información.

A partir del año 2001 se ha venido trabajando en la documentación de los procesos de crítica y codificación, así como en el proceso de digitación. Para este fin se han elaborado informes de los diferentes procesos, control de calidad, rendimientos, etc.

Para la Encuesta de este año se ha iniciado documentando el resto de los procesos, elaborando cronogramas y nombrando los encargados de cada tarea dentro del proceso.

Cuando se genera un cambio en alguna variable este es informado a los usuarios por medio de comunicados de prensa o en los boletines, también en la publicación de los resultados se hace referencias a todos los cambios que se hallan hecho ya sea a la metodología o a las variables.

VI. Integración de Variables

Una vez limpios los datos de la Encuesta de Hogares se procede a crear un archivo tipo DBF que se denomina “**Archivo de Tabulación**”, a este archivo se le deben integrar una serie de variables que no se investigan directamente con el cuestionario sino que se crean a partir de varias preguntas, y en algunos casos de información adicional.

Para crear estas variables adicionales existe un documento denominado “Plan de Creación de Variables” y en el se describen los pasos que se deben seguir para obtener estas variables.

Las variables creadas se agrupan según su tipo, por ejemplo:

- Condición de Actividad
- Ingreso en la Ocupación Principal de los Ocupados Asalariados
- Ingreso en la Ocupación Principal de los Ocupados Independientes
- Ingreso por Otros Conceptos
- Ingreso Primario de los Ocupados Asalariados e Independientes
- Ingreso Total por Trabajo y Otros Rubros
- Condición de Empleo
- Variables de Migración
- Variables de Educación

Además de las variables que se crean para cada una de las personas en este archivo, se integran a este una serie de variables a nivel de hogar, como por ejemplo:

- Tamaño del Hogar
- Clase de Hogar
- Hogares con Ingreso de Asalariados
- Conteo de personas asalariadas
- Hogares con Ingreso de Independientes
- Conteo de personas independientes
- Hogares con Ingreso de No Ocupados
- Ingreso Total del Hogar
- Ingreso per cápita del Hogar
- Ingreso Mínimo del Hogar
- Ingreso Mínimo per. cápita
- Nivel de Pobreza

Los periodos de referencia se mantienen constantes pues la metodología se ha mantenido durante todo el Programa Permanente de Encuestas, esto con el fin de que los datos sean comparables en el tiempo. Los códigos de las variables se mantienen constantes y en caso de que se elimine alguna variable de un año al otro esa variable aparecerá en blanco para ese año.

VII. Arquitectura de las Bases de Datos

Cuando el archivo de tabulación está listo, se generan archivos agregados a nivel de vivienda, a nivel de hogar y a nivel de personas y se brindan a los diferentes usuarios en formatos que puedan ser leídos ya sea en FoxPro, Visual FoxPro, SPSS o Excel, para que puedan ser analizados por los funcionarios que realizan el análisis de los datos.

Con respecto a la documentación de los procesos se está elaborando un Manual de Procedimientos en donde se establecerá la forma en que estos deben ser documentados. En la actualidad esta práctica se ha llevado en forma parcial en el INEC.

VIII. Generación de Tabulados

Para generar los tabulados de la Encuesta de Hogares de Costa Rica se utiliza el módulo **CENTS** del programa **IMPS**. En el caso de esta encuesta se generan tabulados de diferente tipo como por ejemplo:

- Cifras Básicas sobre Fuerza de Trabajo: 6 cuadros
- Cifras Básicas sobre Pobreza e Ingresos: 7 cuadros
- Principales Resultados: 34 cuadros
- Tabulados Generales: 120 cuadros
- Tabulados de la Tercera Edad: 13 cuadros
- Cuestionario de la OIT: 38 cuadros

A parte de esta serie de cuadros que son fijos existe una serie de cuadros adicionales que se generan para atender las solicitudes de usuarios especializados de la encuesta tanto internos como externos.

También cuando las instituciones que solicitan módulos especiales lo requieran se les genera la tabulación de los resultados, de esta forma existen tabulados para los siguientes módulos:

- Vivienda y Servicios
- Telecomunicaciones
- Trabajo Infante Juvenil
- Consumo Aparente de Alimentos

Para el futuro se desea incorporar a este proceso una herramienta de análisis multidimensional con el propósito de que el usuario final tenga a mano toda la información que se produce con la encuesta y pueda el mismo generar sus propias tabulaciones.

IX. Políticas Institucionales de Acceso a las Bases de Datos

Para la ejecución de todos los procesos de la Encuesta de Hogares de Costa Rica existe un cronograma de actividades en donde se consignan las fechas en que deben ejecutarse cada una de las tareas así como los responsables de las mismas.

Dentro de este contexto la entrega de los resultados preliminares se realiza la última semana de octubre y la publicación de los resultados definitivos se hace durante el mes de diciembre.

Los usuarios que deseen tener acceso a la información de la Encuesta de Hogares deben solicitar por medio de una nota a la Coordinadora de la Unidad Técnica de Sistemas e Informática que variables van a necesitar, en que formato lo desean y se es necesario agregar variables adicionales y la forma de cómo crear estas variables. Posteriormente se genera este archivo y se les envía ya sea por correo electrónico o por medio de un medio magnético.

Actualmente se encuentran disponibles en la página del Instituto (www.inec.go.cr) la información sobre los principales resultados de las encuestas de hogares desde el año 1998

En lo que se refiere a las consultas en línea, por medio de un convenio entre el INEC y el Centro Centroamericano de Población de la Universidad de Costa Rica (www.ccp.ucr.ac.cr) se ha puesto a disposición de los usuarios las bases de datos de las Encuestas de Hogares de Costa Rica, que se pueden consultar para la generación de cuadros propios por medio del software Public Data Queris (PDQ Explorer).

También se ha desarrollado una aplicación con el Redatam G-4 para consulta en línea, la cual estará dentro de un mes a disposición en la página Web de la Institución. Ambos sistemas se encuentran disponibles para el procesamiento de la información de los Censos de Población y Vivienda del 2000.

Anexo 1. Plan de Inconsistencias

INSTITUTO NACIONAL DE ESTADÍSTICA Y CENSOS

AREA DE CENSOS Y ENCUESTAS

PLAN DE INCONSISTENCIAS

ENCUESTA DE HOGARES DE PROPÓSITOS MÚLTIPLES

JULIO 2003

San José, Costa Rica

SECCIÓN C: CARACTERÍSTICAS DE LA ACTIVIDAD ECONÓMICA

VARIABLE	DESCRIPCIÓN VARIABLE	REVISIÓN Y ANÁLISIS	CORRECCIÓN
		Nº personas de 12 años o + = Nº registros de Empleo	
NLIN	Nº de Línea	1) Rango: 01 al mayor B1 2) No debe haber dos NLIN iguales. 3) Si B5 = 00 - 11, 98 → NLIN a C39g2 = b* 4) Si B5 = 12 - 97, 99 → NLIN a C39g2 <> b	1) Listar Fuera de rango* 2) Listar si hay 2 ó más iguales* 3) Si B5 = 05 - 11, 98 y alguno de NLIN a C39g2 ≠ b* 4) Si B5 = 12-97, 99 y NLIN a C39g2 <> b
PAREN	Parentesco	1) Rango: 01 - 11 2) Para NLIN = B1, PAREN = B3	1) Listar Fuera de rango* 2) Listar si NLIN ≠ B1 y PARENT ≠ B3 *
SEXO	Sexo	1) Rango: 1 - 2 2) Para NLIN = B1, Sexo = B4	1) Listar Fuera de rango* 2) Listar si NLIN ≠ B1 y SEXO ≠ B4 *
EDAD	Edad	1) Rango: 12 - 97, 99 2) Para NLIN = B1, EDAD = B5	1) Listar Fuera de rango* 2) Listar si NLIN ≠ B1 y EDAD ≠ B5*
INFOR	Informante	1) Rango: 1, 2	1) Fuera de rango
NINFO	Identificación del Informante	1) Rango: 01 al mayor B1, 97, 98, 99, bb 2) Si INFOR = 1 → NINFO = b 3) Si NINFO ≠ NLIN * 4) Si INFOR = 2 → NINFO <= al mayor B1 5) Si INFOR = 2 → NINFO ≠ NLIN 6) Si INFOR = 2 → NINFO ≠ b*	1) Listar si esta fuera de rango. 2) Listar si INFOR = 1 y NINFO ≠ b 3) Listar si NINFO = NLIN * 4) Si INFOR = 2 y NINFO sea mayor al mayor de B1, exepcto 97, 98, 99* 5) Si INFOR = 2 y NINFO = NLIN 6) Si INFOR = 2 y NINFO = b Si INFOR = 2 y NINFO = FR asigne NINFO = A14a Si INFOR = 2 Y NINFO = b, 99 asigne NINFO = A14a*
C1	¿Trabajó la semana pasada?	1) Rango: 1 y 2 2) Si C1 = 1 → C2 a C14 = b 3) Si C1 = 2 → C2 ≠ b	1) Fuera de rango 2) Si C1 = 1 y alguno de C2 a C14 ≠ b* 3) Si C1 = 2 y alguno de C2 a C14 = b*
C2	¿La semana pasada realizó algún trabajo?	1) Rango: 1, 2, b 2) Si C2 = 1 → C3 a C14 = b 3) Si C2 = 2 → C3 ≠ b	1) Fuera de rango 2) Si C2 = 1 y alguno de C3 a C14 ≠ b* 3) Si C2 = 2 y alguno de C3 a C14 = b*
C3	No trabajó la semana pasada pero tiene empleo	1) Rango: 1, 2, b 2) Si C3 = 2 → C4 ≠ b 3) Si C3 = 1 → C4 a C14 = b 4) Si C3 = 1 ó b → C38 a1 a C39 g2 = b	1) Fuera de rango 2) Si C3 = 2 y C4 = b* 3) Si C3 = 1 y alguno de C4 a C14 ≠ b* 4) Si C3 = 1 ó b y alguno C38 a C39g2 ≠ b
C4	Buscó trabajo la semana pasada	1) Rango: 1, 2, b 2) Si C4 = 2 → C5 ≠ b 3) Si C4 = 1 → C5 = b 4) Si C4 = 1 → C6 a C7 ≠ b*	1) Fuera de rango 2) Si C4 = 2 y C5 = b* 3) Si C4 = 1 y C5 ≠ b* 4) Si C4 = 1 y C6 a C7 = b*

Anexo 2. Plan de Creación de Variables

INSTITUTO NACIONAL DE ESTADISTICA Y CENSOS

UNIDAD TECNICA PROGRAMACION, OPERACIONES CENSALES Y ENCUESTAS

PLAN DE CREACION DE VARIABLES

ENCUESTA DE HOGARES DE PROPOSITOS MULTIPLES

San José, Costa Rica

septiembre, 2002

A. Condición de Actividad

Variables utilizadas: C1, C2, C3, C4, C5, C13, C38a1, C39a1, C39b1, C39c1, C39d1, C39e1, C39f1, C39g1

CALCULO	CODIGO	DESCRIPCION
Si C1 o C2 o C3 = 1	01	Ocupado Estándar
Si (C4 o C5 = 1) y C13 = 1 y C38a1 = 1	02	Cesante Ocupado por autoconsumo
Si (C4 o C5 = 1) y C13 = 1 y C38a1 = 2 y ((C39a1 = 1 o 2) o (C39b1 = 4 o 5) o (C39c1 = 1 o 2) o (C39d1 = 4) o (C39e1 = 6) o (C39f1 = 1) o (C39g1 = 3))	03	Cesante Ocupado Marginal
Si (C4 o C5 = 1) y C13 = 1 y C38a1 = 2 y C39a1 = 3 y C39b1 = 6 y C39c1 = 3 y C39d1 = 5 y C39e1 = 7 y C39f1 = 2 y C39g1 = 4	04	Cesante
Si (C4 o C5 = 1) y C13 = 2 y C38a1 = 1	05	Busca por Primera vez Ocupado por Autoconsumo
Si (C4 o C5 = 1) y C13 = 2 y C38a1 = 2 y ((C39a1 = 1 o 2) o (C39b1 = 4 o 5) o (C39c1 = 1 o 2) o (C39d1 = 4) o (C39e1 = 6) o (C39f1 = 1) o (C39g1 = 3))	06	Busca por Primera vez Ocupado Marginal
Si (C4 o C5 = 1) y C13 = 2 y C38a1 = 2 y C39a1 = 3 y C39b1 = 6 y C39c1 = 3 y C39d1 = 5 y C39e1 = 7 y C39f1 = 2 y C39g1 = 4	07	Busca por Primera vez
Si C4 = 2 y C5 = 2 y C38a1 = 1	08	Inactivo Ocupado por autoconsumo
Si C4 = 2 y C5 = 2 y C38a1 = 2 y ((C39a1 = 1 o 2) o (C39b1 = 4 o 5) o (C39c1 = 1 o 2) o (C39d1 = 4) o (C39e1 = 6) o (C39f1 = 1) o (C39g1 = 3))	09	Inactivo Ocupado Marginal
Si C4 = 2 y C5 = 2 y C38a1 = 2 y C39a1 = 3 y C39b1 = 6 y C39c1 = 3 y C39d1 = 5 y C39e1 = 7 y C39f1 = 2 y C39g1 = 4 y C9 = 1 y C10 = 4 o 5	10	Inactivo Desalentado
Si C4 = 2 y C5 = 2 y C38a1 = 2 y C39a1 = 3 y C39b1 = 6 y C39c1 = 3 y C39d1 = 5 y C39e1 = 7 y C39f1 = 2 y C39g1 = 4 y C9 = 1 o 2 y C10 = 1, 2, 3 o 7	11	Inactivo con Deseos de Trabajar

CALCULO	CODIGO	DESCRIPCION
Si C4 = 2 y C5 = 2 y C38a1 = 2 y C39a1 = 3 y C39b1 = 6 y C39c1 = 3 y C39d1 = 5 y C39e1 = 7 y C39f1 = 2 y C39g1 = 4 y C9 = 1 o 2 y C10 = 6	12	Inactivo que ya Encontró Trabajo
Si C4 = 2 y C5 = 2 y C38a1 = 2 y C39a1 = 3 y C39b1 = 6 y C39c1 = 3 y C39d1 = 5 y C39e1 = 7 y C39f1 = 2 y C39g1 = 4 y C9 = 2 y C10 = 4 o 5	13	Inactivo Disponible Estacional
Si C4 = 2 y C5 = 2 y C38a1 = 2 y C39a1 = 3 y C39b1 = 6 y C39c1 = 3 y C39d1 = 5 y C39e1 = 7 y C39f1 = 2 y C39g1 = 4 y C9 = 3	14	Inactivo No Disponible

ASPECTOS METODOLÓGICOS DE LA ENCUESTA DE HOGARES

**OFICINA NACIONAL DE ESTADÍSTICA
CENTRO DE ESTUDIO DE POBLACIÓN Y DESARROLLO (CEPDE)**

CUBA

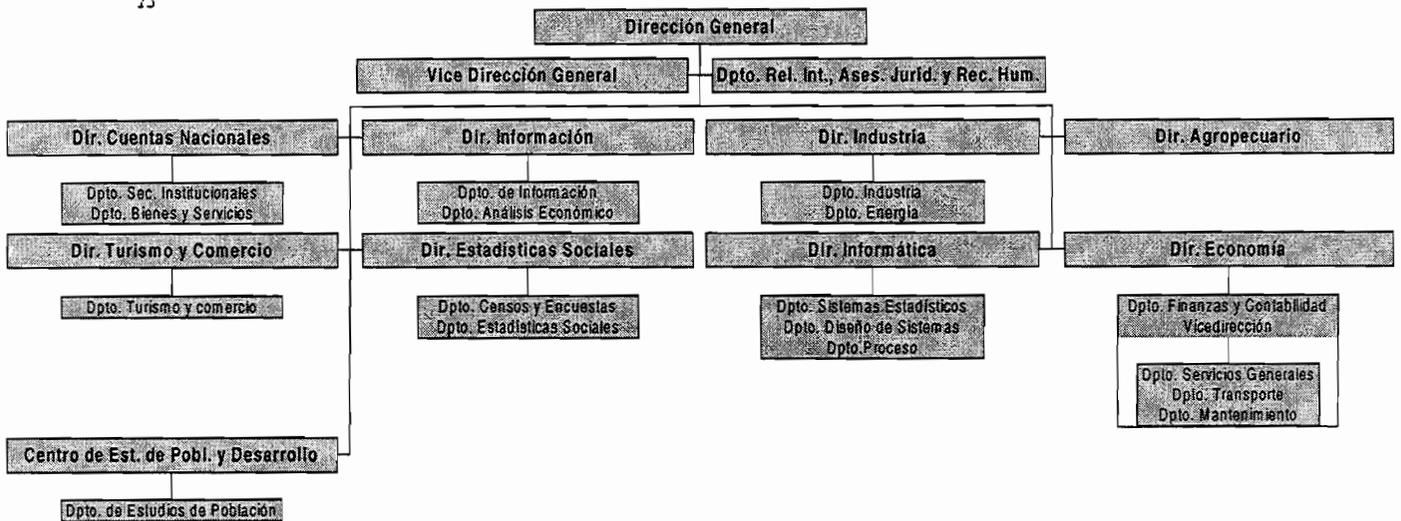
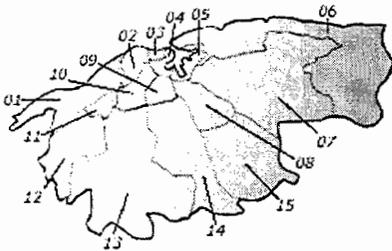
ÍNDICE

	<u>Página</u>
I. Introducción.....	103
1.1 Estructura orgánica de la Oficina Nacional de Estadísticas en Cuba	103
1.2 Estructura orgánica de las Oficinas Territoriales de Estadística.....	103
II. Aspectos Esenciales del Diseño de la Encuesta de Hogares en Cuba	104
2.1 Introducción	104
2.2 Formulario de la encuesta	104
2.3 Definiciones metodológicas básicas	104
2.4 Principios generales	107
2.5 Diseño muestral.....	108
III. Estrategias para la Captura de Datos	109
IV. Procedimientos para Evaluar la Consistencia de la Información	109
V. Diseños Computacionales de las Bases de Datos.....	110
VI. Métodos para la Imputación de Datos	111
VII. Integración de Variables: Gastos, consumo e Ingreso, Desocupados, Subempleados, etc.	111
VIII. Arquitectura de las Bases de Datos	112
IX. Generación de Tabulados	113
X. Políticas Institucionales de acceso a las Bases de Datos.....	113

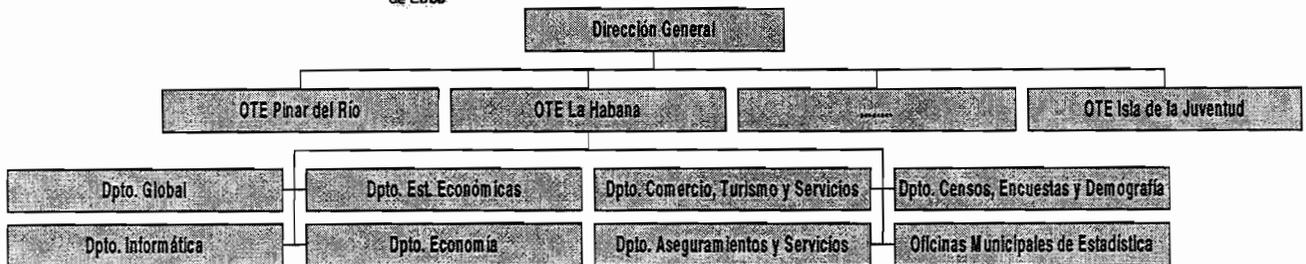
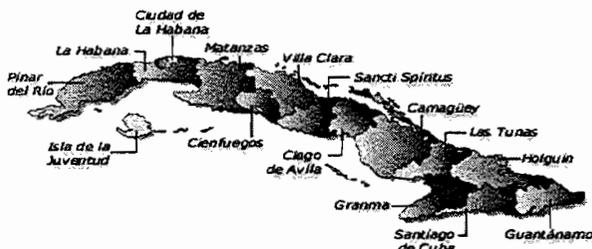
I. Introducción

Para facilitar una mayor comprensión del tema a desarrollar en el 11° Taller se le mostrará primeramente la organización de la Oficina Nacional de Estadísticas en nuestro país y de sus respectivas Oficinas Territoriales en cada provincia, además se detallan algunos aspectos esenciales diseño de la encuesta.

1.1 Estructura orgánica de la Oficina Nacional de Estadísticas en Cuba



1.2 Estructura orgánica de las Oficinas Territoriales de Estadística



II. Aspectos Esenciales del Diseño de la Encuesta de Hogares en Cuba

2.1 Introducción

En las investigaciones relacionadas con la población, es necesario conocer las características socioeconómicas de sus habitantes. Entre éstas se encuentran: las demográficas, educacionales, laborales, niveles y fuentes de ingresos, los gastos y su estructura, la vivienda y su equipamiento y otras.

Para obtener esa información fundamental para el país fue necesario diseñar una encuesta a la que se denominó "Encuesta de Hogares"; la que a partir de módulos temáticos permitirá contar con la tasa de desocupación, fundamental para el Balance de Recursos Laborales y poder estudiar como se ha modificado la situación económica de los hogares, la distribución del ingreso y actualizar informaciones de otros temas según los requerimientos de los diferentes usuarios de la investigación.

2.2 Formulario de la encuesta

Módulo Fijo

Cuestionario que contiene las preguntas sobre las características generales y educacionales y que se llenará en todas las viviendas seleccionadas para participar en la investigación.

Módulo de Ocupación

Cuestionario que contiene las preguntas sobre las características económicas, se llenará en todas las viviendas seleccionadas para participar en la investigación y dentro de éstas a las personas de 15 años y más de edad.

Módulo de Ingresos

Cuestionario donde se reflejarán los ingresos monetarios del núcleo familiar, se llenará en todas las viviendas seleccionadas para participar en la investigación.

Módulo de Gastos

Cuaderno donde se reflejarán los gastos en que ha incurrido el núcleo, durante la semana de referencia, así como mes y doce meses anteriores al mes en que se efectúa la entrevista, su diseño tiene tres versiones, una ampliada, una intermedia y otra reducida, en esta ocasión corresponde levantar la versión reducida, que a diferencia de las restantes se recoge la información mediante entrevista directa al núcleo en la misma visita en que se formulan las preguntas del resto de los módulos.

Este módulo se realizará solamente en aquellas viviendas donde la selección de la muestra se hizo por semanas y se levantará solamente en la parte urbana.

2.3 Definiciones metodológicas básicas

Población a enumerar

Comprenderá todas las personas que integran los núcleos familiares en las viviendas particulares seleccionadas y se enumerará según el concepto de población residente, presente en su vivienda habitual o temporalmente ausente en otro lugar del país o en el exterior.

Vivienda particular

Se considera como vivienda particular todo local o recinto estructuralmente separado e independiente que haya sido construido o adaptado, en todo o en parte, para fines de alojamiento permanente o temporal de personas, con entrada independiente que no tiene que ser directamente con frente a la calle, así como cualquier otra clase de albergue, fijo o móvil, ocupado como lugar de alojamiento.

Si en un local de trabajo existiera en su interior o adosado a la edificación o estructura un recinto

independiente con características de vivienda particular, como son: entrada independiente (por pasillos, jardines, etc), facilidades de cocina, etc., éste sería considerado como tal.

Vivienda colectiva

Es la utilizada como lugar de alojamiento temporal o permanente, por un grupo de personas, usualmente sin vínculo o parentesco, que hacen vida en común por razones de simple convivencia, salud, trabajo, educación, disciplina laboral, religión, etc, ejemplo de ello son los hospitales, hoteles, casas de huéspedes, albergues (estudiantiles) y campamentos (de militares o trabajadores), asilos y otros, que incluyen centros penitenciarios, conventos, etc.

Residentes

Las personas que viven de forma permanente en la vivienda, o sea, aquellas que duermen en la vivienda habitualmente y guardan en ellas sus pertenencias personales.

No obstante la definición anterior no pierde su condición de residencia, en el lugar donde reside el núcleo familiar al cual pertenecen:

- a) Los que por razones de trabajo duermen frecuentemente fuera de la vivienda.
- b) Los que por razones de estudio duermen frecuentemente fuera de la vivienda o en el exterior del país.
- c) Los miembros del cuerpo diplomático y familiares que se encuentran en el exterior del país, desempeñando su cargo u otros residentes en Cuba, temporalmente en el exterior (gestión oficial, artística, culturales, deportivas, enfermos, etc.)
- d) Los que están realizando trabajos u otras funciones en el exterior.
- e) Los ausentes de sus viviendas por encontrarse bajo detención preventiva o bajo medida cautelar.
- f) Los marinos mercantes y los pescadores residentes en el país que se encuentran en el mar territorial o en el extranjero al momento de la Encuesta, así como los demás trabajadores cuyas funciones los precisa salir con cierta regularidad (personal de naves aéreas, etc.)
- g) Los reclusos en centros médicos-asistenciales (hospitales, clínicas, etc.) y que la estancia de los pacientes no es de carácter permanente.

Núcleo familiar u hogar

Se entenderá por núcleo familiar u hogar la persona o grupo de personas con vínculo de parentesco o sin el, pero que viven juntos bajo un mismo régimen económico (con presupuesto común) y que residen en forma habitual o permanente en una vivienda particular o local de trabajo (con características de vivienda particular). En una vivienda puede convivir más de un núcleo familiar, en ese caso se realizará la encuesta a cada núcleo por separado.

Jefe de núcleo

Se entenderá por jefe de núcleo, la persona que se declare como tal o sea considerada así por los demás miembros del núcleo y que generalmente es la persona que aporta la mayor parte de los recursos económicos del mismo.

Nivel educacional más alto terminado completamente

Un nivel se ha terminado completamente cuando se ha aprobado su último año o grado, obteniéndose el diploma o certificado correspondiente como graduado.

Situación económica.

Suele llamarse "tipo de actividad". Es la relación que existe entre cada persona y su actividad económica. Se determina mediante una clasificación general de la población que permite establecer si la persona es o no económicamente activa. La determinación de la condición de actividad está íntimamente relacionada con los límites de edad establecidos y con el período de referencia fijado para la investigación. El límite mínimo de edad a partir del cual se clasificará la población según su actividad será de 15 años. La situación laboral o condición de actividad permite clasificar a la población en dos grupos:

I. Población económicamente activa

La población económicamente activa abarca a todas las personas que reúnen los requisitos para ser incluidas entre las personas empleadas o las personas desempleadas.

Es decir, se refiere a los ocupados (de 15 años y más) y a los desocupados (hombres de 17 a 59 años y mujeres de 17 a 54 años), que no trabajan y están buscando empleo durante el período fijado para la investigación.

Población ocupada

Personas de 15 años y más que en el período de referencia "estaban trabajando" es decir laboraron no menos de 8 horas semanales en una ocupación del sector estatal, cooperativo, privado u otros, tanto dentro del territorio nacional como en el extranjero. Incluye trabajadores que no laboraron en el período de referencia por: disfrutar de vacaciones reglamentarias, enfermedad, accidente, licencia, capacitándose en el país o en el extranjero, movilizado, afectaciones por materias primas, combustibles, etc.

Población desocupada

Personas en edad laboral es decir, hombres de 17 a 59 años y mujeres de 17 a 54 años, que no trabajaron o lo hicieron por menos de 8 horas, que no tenían vínculo laboral estable, y que en el período de referencia buscaban empleo por primera vez o porque lo habían perdido.

A los fines de esta investigación, se considerarán como desocupados los disponibles con garantía salarial y los disponibles con subsidio.

II. Población no económicamente activa

Comprende aquellas personas de 15 años y más, de ambos sexos, que no tienen vínculo laboral alguno con la producción de bienes y prestación de servicios y no buscan empleo.

Este grupo comprende: estudiantes, pensionados por edad, invalidez total o parcial y otros, rentistas, personas que reciben ayuda económica, personas que se dedican a quehaceres del hogar, incapacitados para el trabajo, personas fuera de la edad laboral que no trabajan ni estudian, personas capacitándose y personas en otra situación.

Personas desalentadas.

Son aquellas personas que desean un trabajo y que actualmente están dispuestas a trabajar, pero han abandonado toda búsqueda activa porque creen que no lo encontrarán. Los trabajadores desalentados al igual que los desocupados representan recursos de mano de obra no utilizados y es necesario tener información sobre ellos para una evaluación completa de la situación del mercado de trabajo.

Ingresos monetarios.

Comprende todas las formas de ingreso en efectivo percibido por la población. Las modalidades en que pueden percibirse estos ingresos son: salarios, premios, dietas y otros tipos de remuneraciones, ingresos de campesinos, estipendios, prestaciones a corto plazo (maternidad, enfermedad y accidentes) y a largo

plazo (jubilaciones y pensiones), subvenciones, compensaciones a antiguos propietarios, préstamos estatales, ingresos de transportistas privados, ventas de productos y servicios productivos y otros ingresos tales como: donaciones, intereses por depósitos bancarios, subsidios e indemnizaciones a agricultores pequeños, arrendamiento de fincas, etc.

Gastos

Es el valor de los gastos o salidas monetarias en que incurre el núcleo familiar. La naturaleza de los mismos puede ser: alimentación, ropa calzado, transporte, diversiones, pago de la vivienda, electricidad, combustible, muebles, equipos eléctricos y todo tipo de gastos, tanto del hogar como de carácter personal, incluyendo en los que se incurre por pagos de actividades privadas (consultas médicas, dentistas, modistas, etc.), así como los gastos ocasionados por la producción propia y otros no vinculados al consumo como pago de cuotas a organizaciones, pensiones alimenticias, multas, intereses por préstamos, etc.

2.4 Principios generales

Período de enumeración

La enumeración se efectuará durante los meses de marzo a mayo teniendo en cuenta lo siguiente:

Las viviendas seleccionadas para participar en el primer y segundo mes solamente tendrán información en los Módulos Fijo, Ocupación e Ingresos.

Las viviendas seleccionadas para participar en el tercer mes pueden tener las siguientes características:

Que la muestra esté dispuesta por semanas, en ese caso a las viviendas contenidas en ella se le recogerá información en los Módulos Fijo, Ocupación, Ingresos y Gastos, aunque puede suceder que en aquellos casos en que las personas no deseen dar información de los gastos solamente aparezcan el resto de los módulos con datos

Que la muestra no esté dispuesta por semana, en ese caso los únicos módulos que deberán aparecer con información en esas viviendas serán Módulo Fijo, Ocupación e Ingresos.

Universo y alcance

La encuesta abarca a toda la población residente en viviendas particulares y locales de trabajo (con viviendas particulares) de la parte urbana y rural del país.

Se excluyen los residentes en colectividades, ingresados permanentemente en hospitales y sanatorios.

Método de enumeración

El método de enumeración consistirá en la entrevista directa a todos los integrantes del núcleo. Si algunas de las personas a entrevistar no se encuentra en ese momento en la vivienda, se solicitarán los datos de las mismas a otra persona mayor de edad que pueda suministrar la información.

Aspectos a investigar

- **OCUPACION:** Se investigará la Ocupación principal.
- **DESOCUPACION:** Gestiones que realiza para conseguir un empleo, dificultades para encontrarlo, tiempo de búsqueda, personas que dependen de él, causas que motivaron la pérdida del empleo, si es desocupado por primera vez o si es por perder un empleo anterior.
- **POBLACION NO ECONOMICAMENTE ACTIVA:** Dentro de esta se investigan los Pensionados, Estudiantes, Quehaceres del Hogar, No realiza ninguna actividad y Otra situación, así como las personas desalentadas.

Gastos monetarios.

Ingresos monetarios.

Período de referencia

- ◆ *Doce meses* anteriores al mes en que se realiza la Encuesta para los ingresos y los gastos que en los referidos módulos se indican con tal periodicidad.
- ◆ *Mes anterior* al mes en que se realiza la Encuesta, para las informaciones que en los diferentes módulos se indican con tal periodicidad.
- ◆ *Semana anterior* al "día de la Encuesta", (domingo a sábado anteriores) los datos contenidos en los módulos de ocupación y gastos se indican con tal periodicidad..
- ◆ *Día de la Encuesta*, los datos que se solicitan en las preguntas 1 a 5 (Características Generales y Educativas).

2.5 Diseño muestral

Marco Muestral

Se seleccionó la muestra del Diseño Muestral General o Muestra Maestra del Sistema de Encuesta de Hogares el cual tuvo como Marco Muestral Maestro el elaborado por la ONE para la realización de encuestas de hogares en 1995 y que fue actualizado en 1999.

Método de Muestreo

El método de muestreo que se utilizó corresponde con un **muestreo por conglomerados trietápico**, con estratificación de las unidades primarias. Las provincias y dentro de ellas los municipios constituyeron los estratos.

Rotación de la muestra

En este año 2002 se mantuvo, como en el 2001, la unificación en una sola encuesta de las dos encuestas continuas (Encuesta Nacional de Ocupación y Encuesta sobre la Situación Económica de los Hogares) con el nombre de Encuesta de Hogares y en un mismo período de levantamiento.

En el año 2002 se mantuvo la rotación del 50% de las viviendas en todos los módulos.

Tamaño de la Muestra

El tamaño de la muestra se determinó de manera tal que permitiera obtener resultados de los indicadores más importantes al nivel de provincias y nación, con una caída de muestra esperada promedio de un 10% y con un nivel de error inferior al 5%. Todo esto permitirá hacer a posteriori un análisis de indicadores con cierto grado de desagregación. El tamaño de muestra para Cuba fue de 39 303 viviendas. El tamaño de la submuestra de viviendas que participaron en el módulo de gasto fue de 7904.

Estimación

Los estratos de ponderación fueron los municipios parte urbana y rural dentro de cada provincia. Se aplicaron estimadores de razón que utilizaron como variable auxiliar el **Total de Población** que ofrecen los Cálculos Demográficos del país.

Para las variables de Ocupación e Ingresos se aplicó además una post-estratificación por sexo y grupos de edad (15-29, 30-54, 55 y más para las Mujeres y 15-29, 30-59, 60 y más para los Hombres)

III. Estrategias para la Captura de Datos

La captación de datos se realiza descentralizada en los Centros de Cálculos Provinciales (15 en total) y en algunas provincias se realizan en las Oficinas Municipales de Estadísticas.

Los cuestionarios son llenados en el terreno por el enumerador, posteriormente llevan un proceso de validación de consistencia de la información a diferentes niveles (Supervisor, Oficinista Municipal, Oficinista Provincial) culminado este paso se digitaliza en los centros de cálculos por un personal especializado en esas funciones.

Las preguntas que requieren de codificación a posterior en las oficinas se ajustan a los codificadores establecidos, ejemplos de codificadores: empresa, cargos, división política administrativa, etc. Las preguntas de la encuesta son digitalizadas directamente según el código establecido por los metodólogos que participan en la investigación.

IV. Procedimientos para Evaluar la Consistencia de la Información

Los datos se validan en un primer momento en el centro de cálculo provincial al digitarse la información empleando los sistemas de carga de la información elaborados por los especialistas en Análisis de Sistemas de la ONE. La validación es sencilla pero con alta importancia para no demorar innecesariamente la carga de datos. La información digitalizada es verificada al 100% emitiéndose un listado de errores cuando exista inconsistencia para su comprobación a posteriori con el cuestionario.

Para cada pregunta se realiza un chequeo de rango y de secuencias lógicas establecidas por saltos de preguntas contemplados en los programas elaborados.

La captura de datos se realiza con un diseño de fichero donde se conforma la información por tipos de registros según su temática

- ◆ Registro de viviendas en el cual se registran los datos de la vivienda
- ◆ Registro de personas en el cual se registran los datos referidos a la persona
- ◆ Registro de gastos en el cual se registran los gastos incurridos por hogar en la vivienda

Los ficheros provinciales son enviados al departamento de Análisis de la ONE por e-mail y la unificación de todas conforman la del país. El tiempo que demora este proceso varía de acuerdo a la complejidad de la encuesta, la cantidad de información a digitalizar y la cantidad de personal asignado para la digitalización.

En nuestra encuesta no es necesaria la imputación automática para la información faltante porque en su diseño toda pregunta está asociada a un código de respuesta o una codificación a posteriores en algunas preguntas, además de que un requerimiento muy importante establecido en la validación de los cuestionarios por los supervisores y oficinistas es que ninguna pregunta carezca de información para darse por aceptado el formulario.

Los trabajadores de la dirección de informática entregan al especialista responsable de la encuesta del departamento de Estadísticas Matemáticas Aplicadas una base de datos en FoxPro para cada tipo de registro (personas, viviendas y gastos) en donde se realiza una última validación que contempla el universo de las unidades maestras por estrato muestral, universo de personas, núcleos y viviendas que participaron en la investigación chequeándose con los reportes enviados a nuestro departamento por la provincia (cobertura de la muestra). En particular las viviendas que intervienen en el Módulo de Gasto son validadas si participan según la distribución por semanas establecida. Por último se valida de nuevo el rango de cada pregunta, los cruces de indicadores y las variables agregadas (nuevas) que se forman por programación de las variables originales.

Se interactúa con el personal responsable en las Oficinas Provinciales de Estadísticas para la aclaración de las inconsistencias encontradas. Este proceso se continúa hasta que no se detecten más errores y quede clasificada la base de datos limpia de errores para emitir las tablas de salida de la información y por consiguiente la realización de análisis de los datos. Este proceso de interacción con las provincias demora a lo sumo 15 días. La emisión de las tablas de salida se realiza a los 5 meses de iniciado el levantamiento.

V. Diseños Computacionales de las Bases de Datos

En la elaboración del **diseño de fichero** se contempla los siguientes aspectos

- ❖ *Establecer el tipo de registro*
- ❖ *Creación del Diccionario de Datos*
- ❖ *Carga y Validación*
- ❖ *Creación del fichero rectangular con longitud variable*
- ❖ *Edición de las tablas*

Para la captura de la información y la validación en ese momento se utilizan diferentes software, tales como PCEDIT de las Naciones Unidas (Dpto. de Desarrollo de software), DD, CENTRY del IMPS los cuales fueron elaborados por el Buró de Censos de los Estados Unidos. Después de culminada la captura y creados los ficheros se usa el CONCOR del IMPS para ejecutar una nueva validación y programas elaborados en CLIPPER por los especialistas en Sistema Automatizado de la ONE

En el departamento de Estadística Matemática al recibir la base de datos unificadas para la Nación se utilizan diferentes software según la finalidad.

Validación: Programas elaborados por el especialista a cargo de la investigación en FoxPro.

Imputación del Factor de Expansión: Software elaborado la dirección de informática de la ONE

Procesamiento:

- Para la emisión de tablas y la formación de cluster se utiliza el SPSS
- Para el cálculo de las estimaciones, los errores muestrales, intervalos de confianza, coeficientes de variación, efecto de diseños, correlaciones, regresión logística, X^2 , comparación de proporciones en dos momentos diferentes, etc. se utiliza el software STATA.
- El PCCARP se utiliza para el cálculo de los errores muestrales, el efecto de diseños, X^2 , regresión múltiple.

Los diseños muestrales empleados en las encuestas de hogares corresponden a diseños muestrales complejos, requiriendo tener en cuenta este aspecto para las estimaciones de los indicadores con sus respectivos errores absolutos y relativos y para los correspondientes análisis estadísticos en que son empleadas las herramientas matemáticas. Está aspecto mencionado anteriormente obliga a la utilización solamente de algunos softwares que son confeccionados para diseños muestrales complejos.

La versión del SPSS que disponemos es la número 10 pero no en su versión completa al faltarle los módulos del cálculo de los errores para Diseños Muestrales Complejos, cálculo del tamaño de muestra entre otros módulos, motivo por el cual se utiliza solamente para emitir tabulados sencillos que no requieran de los errores muestrales, tampoco es de utilidad para realizar los análisis de datos.

VI. Métodos para la Imputación de Datos

Como hemos mencionado anteriormente la *No respuesta Total o Parcial* se trata en primer lugar de recuperar esta información contacto por medio del enumerados a la persona involucrada y por consiguiente después digitalizarla, luego no ha sido necesario realizar una imputación automática por información faltante.

VII. Integración de Variables: Gastos, Consumo e Ingreso, Desocupados, Subempleados, etc.

Se realiza lo que se llama preparación de información donde se crean variables que se capturaron en el formulario. Ejemplo de ello son la formación de indicadores tales como Ocupados, Desocupados, Población Económicamente Activa, Ingresos Totales, Ingreso Per cápita, etc.

La definición de estos indicadores es explícitamente explicada y redactadas en un documento para los especialistas de sistemas, después de su creación son validadas con las variables originales para determinar inconsistencias y solucionarlas, además de vincular los indicadores que están relacionados.

Ejemplo:

MODULO DE OCUPACION (PERSONAS DE 15 AÑOS O MAS DE EDAD)		SEM	MES
No. D E O R D E N	1. ¿Trabajó la semana del ___ al ___, al menos 8 horas por un salario, beneficio o ganancia familiar? SI.....1 Pase a 4 NO.....3 Pase a 2	2. ¿Aunque no trabajó tenía algún empleo del cual estuvo ausente? SI.....1 Pase a 3 NO.....3 Pase a 7 Pase a 7: personas en edad laboral, mujeres de 17 a 54 años y hombres de 17 a 59 años. FUERA DE ESTAS EDADES PASE A 14.	3. ¿Por qué motivo no trabajó esa semana? Enfermedad o accidente.....1 Licencia.....2 Vacaciones.....3 Capacitándose.....4 Movilizado.....5 Afectaciones por materias primas, combustibles, etc.....6 Otros.....7 Pase a 4
	1 3	1 3	1 2 3 4 5 6 7

Sí $P1=1$ ó $P2=1$ entonces $Ocupados=1$

DESOCUPADOS (PERSONAS EN EDAD LABORAL)			
No. D E O R D E N	7. ¿Buscó algún tipo de empleo durante las cuatro semanas anteriores? SI.....1 Pase a 8 NO.....3 Pase a 14	8. ¿Qué gestiones realizó para conseguir el empleo? Consultó amigos o Parientes.....1 Consultó directo a empleadores.....2 Consultó al órgano de trabajo.....3 Otras.....4 (SI NO HIZO GESTIONES REVISE LA PREGUNTA 7 Y DETERMINE LA ALTERNATIVA CORRECTA, PARA BUSCAR EMPLEO DEBE HABER HECHO ALGUNA GESTION) Pase 9	9. ¿Qué dificultades presenta para conseguir el empleo? No siente motivación por las ofertas de empleo.....1 No encuentra oferta de empleo.....2 Exigían años de experiencia.....3 Otras.....4 Pase a 10
	1 3	1 2 3 4	1 2 3 4

Sí $P1=3$ y $P2=3$ y $P7=1$ entonces $Desocupados=1$

$PEA = Ocupados + Desocupados$

$Población de 15 años y más = PEA + PNEA$

Gastos son recalculados y llevados a información mensual

VIII. Arquitectura de las Bases de Datos

Después de culminada validación por el CONCOR del IMPS y obtenida una base limpia de errores, se confeccionan programas en FOXPRO por los especialistas en Sistema Automatizado de la ONE que contienen toda la información digitalizada y enviados al Departamento de Estadísticas Matemática Aplicada en un fichero que es nuevamente validado y que se le aplica el software FACTORI para imputar el factor a nivel de persona, gasto y núcleo.

En FOXPRO se crean nuevos campos teniendo en cuenta el plan de tabulación

Para agilizar el proceso de edición se crean en FOXPRO diferentes ficheros según tipo de información solicitada en el plan de tabulaciones, por lo que se pueden crear ficheros de información de núcleos y personas, gastos y personas, personas, núcleos, gastos, etc.

Estos nuevos ficheros posteriormente son copiados tipo SDF para ser explotados por el CENTS del IMPS.

Al departamento de Estadísticas Matemática Aplicada se les entregan tres bases independientes en FOXPRO por tipo de registro, las cuales contienen la siguiente información

- **Personas:** Contiene el módulo fijo, módulo de ocupación y el modulo de ingreso añadiéndose además el ingreso per cápita de cada persona, las variables del diseño (estrato, cluster, factor) y otros indicadores nuevos.
- **Núcleos:** Contiene por núcleo los totales de ingresos recogidos y posteriormente recalculados a información mensual tomando como base la información registrada en el módulo de ingreso, los totales de personas y algunos datos sociodemográficos del jefe de núcleo, tales como: (sexo, edad, nivel educacional, cantidad de pensionados, cantidad de trabajadores estatales), además se le incorpora las variables del diseño (estrato, cluster, factor) y otros indicadores nuevos.
- **Gastos:** Los gastos son identificados por un código único y se dividen en grupos y por período de referencia del gasto La información es referida al gasto incurrido por el núcleo familiar y es reprocesado para obtener la información mensual, se digitaliza los pesos y centavos en moneda nacional y en divisas. Para esclarecer la recogida de la información le mostramos una tabla resumen por grupo y periodicidad.

RESUMEN

GRUPO	CODIGO
PERIODICIDAD SEMANAL	
ALIMENTACION EN EL HOGAR	01
INFUSIONES BEBIDAS ALCOHOLICAS Y NO ALCOHOLICAS	02
TABACOS, CIGARROS Y FORFOROS	03
TRANSPORTE	04
COMBUSTIBLES DOMESTICOS	05
GAS	06
ALIMENTACION FUERA DEL HOGAR	07
PERIODICIDAD MENSUAL	
ELECTRICIDAD, AGUA Y ALCANTARILLADO	08
JABONES, DETERGENTE Y PASTA DENTAL	09
SERVICIOS DE TRANSPORTE Y COMUNICACIONES	10
OTROS BIENES Y SERVICIOS	11
GASTOS AJENOS AL CONSUMO	12
PERIODICIDAD ANUAL	
OTROS BIENES Y SERVICIOS	13
GASTOS AJENOS AL CONSUMO	14

CODIGOS LUGARES DE ADQUISICION O PAGO

LUGAR DE ADQUISICION O PAGO	CODIGO
ALIMENTACION EN EL HOGAR Y RESTO DE BIENES Y SERVICIOS	
ESTABLECIMIENTOS ESTATALES (INCLUYE CENTRO DE TRABAJO	1
MERCADO AGROPECUARIO	2
OTROS VENDEDORES Y PERSONAS PARTICULAES	3
ALIMENTACION FUERA DEL HOGAR	
CAFETERIAS RESTAURANTES Y OTROS ESTABLECIMIENTOS ESTATALES (INCLUYE COMEDOR CENTRO DE TRABAJO Y ESTUDIO)	4
CEFETERIAS, RESTAURANTES Y OTROS ESTABLECIMIENTOS PRIVADOS Y PERSONAS PARTICULARES	5

LUGARES DE ADQUISICION ADMISIBLES PARA CADA GRUPO

GRUPO	LUGARES DE ADQUISICION
01	1,2 y 3
02	1 y 3
03	1 y 3
04	1 y 3
05	1 y 3
06	1
04	4 y 5
08	1 y 3
09	1 y 3
10	1 y 3
11	1 y 3
12	1 y 3
13	1 y 3
14	1 y 3

Para la emisión de tabulados se analizan los que estén relacionados por el tipo de información que recoge y para cumplimentar este propósito son creadas bases de datos independientes con la información imprescindible.

A su vez las bases en FoxPro son enviadas al especialista responsable de esa investigación del departamento de Estadística Matemática Aplicada, las cuales son importadas por los paquetes estadísticos SPSS, STATA.

Para la utilización del paquete PCCARP se exporta la base de datos recibida en un fichero texto delimitado por blancos.

IX. Generación de Tabulados

Se realizan los tabulados por CENTS del IMPS, SPSS, STATA y PCCARP

X. Políticas Institucionales de Acceso a las Bases de Datos

Cuando es conformada la investigación por el equipo multidisciplinario en que intervienen psicólogos, sociólogos, analistas de sistema, muestristas, entre otros; se elabora un cronograma de trabajo discutido por todos los especialistas que intervienen en la investigación. Los usuarios realizan la solicitud de los tabulados a nuestra ONE y son entregados en el tiempo previsto.



ENCUESTA DE HOGARES – SITUACIÓN ACTUAL

CARLOS OBANDO

INSTITUTO NACIONAL DE ESTADÍSTICA Y CENSO (INEC)

ECUADOR

ÍNDICE

	<u>Página</u>
I. Estrategias para la Captura de Datos y Procedimientos para Evaluar la Consistencia de la Información	117
1. Encuesta de Empleo Urbano - ENEMDU	117
2. Encuesta de Condiciones de Vida - ECV	117
3. Encuesta de Ingresos y Gastos - ENIGHU	118
4. Encuesta de Indicadores de la Niñez y los Hogares	118
II. Diseños Computacionales de las Bases de Datos	119
1. Encuesta de Empleo y Desempleo	119
2. Encuesta de Condiciones de Vida	119
3. Encuesta de Ingresos y Gastos	120
4. Encuesta de Indicadores de la Niñez y los Hogares	120
III. Métodos de Imputación de Bases de Datos	120
IV. Integración de Variables: Gasto, Consumo e Ingreso. Desocupados, Subempleados, etc. .	121
V. Arquitectura de la Base de Datos	121
1. Encuesta de Empleo	121
2. Encuesta de Condiciones de Vida	121
3. Encuesta de Ingresos y Gastos	122
4. Encuesta de Indicadores de la Niñez y los Hogares	122
VI. Generación de Tabulados	122
VII. Políticas Institucionales de Acceso a las Bases de Datos	122

I. Estrategias para la Captura de Datos y Procedimientos para Evaluar la Consistencia de la Información.

1. Encuesta de Empleo Urbano - ENEMDU

- Esta encuesta se viene realizando cada mes de diciembre desde 1990 hasta la fecha.
- La captura de datos se realiza de manera descentralizada en cuatro oficinas regionales, una vez que los formularios de las encuestas han sido previamente diligenciados en el campo, criticados y codificados.
- El sistema de ingreso de datos está implementado para funcionar en ambiente multiusuario, capaz de una vez finalizada la digitación enviar la base de datos a la dirección central para la consolidación de la base de datos nacional.
- Antes de ser enviadas las base de datos regionales, estas se someten al proceso automático de consistencia, validación y cobertura, a fin de tener una base lo mas depurada posible; de tal manera que si existen inconsistencias, estas son corregidas en base a los formularios de las encuestas por parte de un grupo de personas contratadas que realizan el trabajo de crítica y codificación.
- La codificación se realiza de manera manual, en aquellas preguntas que hacen referencia a la rama de actividad y grupo de ocupación, por parte de un grupo de personas contratadas llamadas críticos codificadores.
- El ingreso de datos se realiza las tres primeras semanas del mes de diciembre y de manera paralela al levantamiento de la información.
- Para una muestra nacional de 715 sectores de 12 viviendas cada uno, se utilizan un total de 12 digitadores.

2. Encuesta de Condiciones de Vida - ECV

- En Ecuador se han realizado hasta la fecha cuatro encuestas de este tipo (1994, 1995, 1998 y 1999)
- Para cada encuesta en su momento se desarrollaron sistemas de ingreso de datos diferentes debido a que de igual forma se dieron cambios sustanciales al formulario de hogares que se utilizó.
- La digitación se realizó en el campo, de tal manera que el equipo de trabajo encargado de realizar el trabajo de campo, estaba conformado por un supervisor, tres encuestadores y un digitador.
- Debido a la metodología de recolección de información que consistía en realizar la investigación a un hogar en dos visitas, se realizó de igual forma sistemas de ingreso de datos modulares.
- El digitador debía ingresar la información de 6 sectores, cada uno de 9 viviendas, en un tiempo de 20 días de trabajo mas cuatro de descanso, que era la carga de trabajo que tenía el grupo.
- La consistencia, validación y cobertura, se realizaba día a día utilizando un módulo adicional al sistema de ingreso de datos, que se desarrollo en FOXPRO (ver 2.6). También la corrección de inconsistencias se realizaba día a día de manera paralela al levantamiento de la información.

- De manera periódica y especialmente cuando se sucedían cambios de localidad el supervisor enviaba los archivos de datos a las oficinas regionales, para luego el coordinador regional de la encuesta entregar estos a planta central.
- En la oficinas de planta central se disponía de un grupo de cuatro digitadores encargados de redigitar el 50% de la información recibida de las direcciones regionales, previa la consolidación de la base de datos nacional.
- El digitador a más de la tarea de digitación, tenía el trabajo de criticar la información y codificar de manera manual todas aquellas preguntas tales como rama de actividad, grupo de ocupación, cursos de capacitación, unidades de medida, etc.

3. Encuesta de Ingresos y Gastos - ENIGHU

- Esta encuesta no tiene una periodicidad bien establecida, sin embargo de ello hoy en día se está llevando a cabo la ENIGHU 2003, que tiene una duración de un año. Esta encuesta tiene el objetivo de investigar en gran medida los ingresos y gastos que tienen los hogares ecuatorianos en una economía dolarizada, con la finalidad de realizar un análisis del presupuesto familiar y de paso también determinar la nueva canasta del Índice de Precios al Consumidor Urbano.
- El ingreso de datos se realiza de manera centralizada en las oficinas de la planta central del INEC, utilizando una aplicación para ambiente multiusuario, de tal manera que los controles de validación y cobertura se realizan también en planta central.
- La crítica y codificación de los formularios se realiza en las oficinas regionales previo el envío de estos a planta central para su revisión e ingreso de datos, por lo que en planta central se dispone de 10 digitadores y 10 revisores.
- En caso de encontrar errores sistemáticos o frecuentes, se dan las directrices correspondientes utilizando el INTERNET.
- Tomando en cuenta que la muestra es de 11256 viviendas distribuidas en todo el año, un digitador tiene una carga mínima de 6 viviendas diarias.
- A mas de los 10 digitadores existen 2 digitadores revisores, quienes verifican el 50% de la información digitada.
- Al final de cada periodo (un mes de encuesta) se realizan los procesos de consistencia, validación y cobertura en forma automática, de tal manera que una vez que se hayan detectado los errores sean estos de digitación o de consistencia de información por parte de un grupo de revisores, se escoge a dos digitadores para realizar las correcciones respectivas.

4. Encuesta de Indicadores de la Niñez y los Hogares

- Esta se realizó en noviembre y diciembre el año 2000 juntamente con la Encuesta de Empleo, cuyo objetivo fue el de recabar información respecto a variables que muy poco se habían investigado, tales como el empleo infantil, la migración internacional, seguridad ciudadana, salud infantil, etc., que permitan determinar indicadores importantes respecto de la niñez y los hogares ecuatorianos.
- El ingreso de datos se realizó utilizando una aplicación desarrollada en IMPS. Esta actividad fue realizada en las respectivas direcciones regionales que tiene el INEC, de tal manera que una vez realizado el ingreso y posteriormente la validación y cobertura, se realizó la consolidación de la base de datos nacional en las oficinas de la planta central.

- Al igual que la encuesta de condiciones de vida, la consistencia, cobertura y parte de la validación se realizó utilizando un módulo desarrollado en FOXPRO (ver 2.6).
- Como esta encuesta se levantó juntamente con la de Empleo, la muestra que se utilizó fue la misma, pero como el volumen de información era mayor se necesitó de casi el doble de digitadores que comúnmente se utilizan para la encuesta de Empleo.

II. Diseños Computacionales de las Bases de Datos

1. Encuesta de Empleo y Desempleo

- Desde 1990 hasta 1999 el ingreso de datos se lo ha realizado utilizando una aplicación desarrollada en FOXPRO, la misma que año a año se le realizó ajustes y depuraciones de tal manera que en el último año que se utilizó se llegó a un nivel de eficiencia por parte de la aplicación de casi un 100%. A partir del 2000 se rediseñó el formulario y por ende el sistema de ingreso de datos, el cual se elaboró en VISUAL FOXPRO (ver 5.0)
- La validación, cobertura y consistencia de la información se la realiza a través de opciones implementadas en el mismo sistema de ingreso de datos.
- En varios casos debido a la falta de recursos económicos el parque informático que se disponen en las direcciones regionales no es de lo último que ofrece el mercado, por lo que se ha visto siempre en la necesidad de utilizar un software no tan exigente en recursos, de ahí que se ha optado por la política de “ir del problema al software” es decir que en base a las circunstancias y necesidades se escoge el software más apropiado, que en este caso ha sido el FOX.
- En cuanto se refiere a la explotación de las bases de datos, esto se lo realiza utilizando el paquete estadístico SPSS, tal es así que se viene trabajando desde la versión 3.1 hasta hoy en día trabajar con la versión 10.
- Las base de datos en términos generales ha tenido el mismo diseño, puesto que las entidades que se manejan son prácticamente las mismas, debido a que las unidades de estudio son los hogares y al interior de estos son las personas. Es importante hacer notar el hecho que se han mantenido los mismos nombres y códigos para las variables consideradas importantes o medulares para la encuesta, de tal manera que se mantiene un estándar que permite trabajar fácilmente con las bases de datos de cualquier año.

2. Encuesta de Condiciones de Vida

- Por la naturaleza y particularidades de este tipo de encuesta, el ingreso de datos se realizó utilizando aplicaciones desarrolladas en IMPS, una por cada encuesta.
- La validación de la información se realizó utilizando programas desarrollados en CONCOR (módulo del IMPS), de tal manera que al finalizar una carga de trabajo el digitador ejecutaba dichos programas para obtener un reporte de errores que debían ser corregidos ese momento o al siguiente día dependiendo del tipo de error.
- La cobertura y consistencia de la información se realizó utilizando una aplicación desarrollada en FOX (ver 6.0) la misma que estaba integrada a través de un software utilitario con la aplicación de IMPS, de tal manera que el digitador vía opciones podía interactuar con cualquiera de las dos de acuerdo a su necesidad.

- El por qué se utilizó IMPS, fue por su ahorro de tiempo en cuanto a trabajo de programación, por permitir desarrollar aplicaciones que manejan longitudes de registros , además por su facilidad de manejo que permitió un ingreso en el campo efectivo.
- Una vez consolidada la base de datos nacional en formato ASCII, se procedió a exportar a formato SPSS a fin de realizar el procesamiento y/o explotación de las bases de datos.
- El diseño de las bases de datos para este tipo de encuestas están diseñadas tomando como marco referencial la filosofía que impone el IMPS, cual es trabajar con cuestionarios y tipos de registro.

3. Encuesta de Ingresos y Gastos

- Actualmente el ingreso de datos de esta encuesta se realiza a través de una aplicación desarrollada en VISUAL FOX (ver 5.0), que permite realizar un ingreso en ambiente multiusuario.
- La consistencia, cobertura y validación de la información se realiza de manera automática, utilizando la aplicación que se utiliza para el ingreso de datos, ya que esta contempla opciones para el efecto.
- La base de datos responde a un diseño Entidad-Relación que está de acuerdo a cómo se estructura el formulario.
- La explotación de la base se realizará en SPSS.
- El por qué se decidió utilizar el VISUAL FOX, fue porque se tiene una muy buena experiencia con esta herramienta y por el hecho que se planificó implementar una aplicación que permita realizar el mayor número de controles a nivel de rangos, flujos y consistencia de información, en el momento mismo del ingreso, a fin de obtener una base de datos lo mas depurada posible.

4. Encuesta de Indicadores de la Niñez y los Hogares

- Para el ingreso de datos se utilizó una aplicación desarrollada en IMPS, de tal manera que el diseño de la base de datos se realizó en base a la filosofía que impone esta herramienta.
- Al igual que la encuesta de condiciones de vida se implementó un módulo desarrollado en FOXPRO (ver 6.0) que permitió realizar, la consistencia, cobertura y validación de la información de los archivos ASCII, más comunmente llamados en IMPS como BATCHS.
- En este caso, por el número de variables que tenía la encuesta, la urgencia de implementar una aplicación para el ingreso de datos y aspectos de carácter logístico y administrativo, hicieron que se decida realizar el ingreso con una aplicación desarrollada en IMPS.
- La explotación de las bases de datos al igual que en todas las encuestas de hogares, se realizó en SPSS.

III. Métodos de Imputación de Bases de Datos

Para el efecto en todas las encuestas de hogares se adoptan simbologías y códigos que permiten estandarizar situaciones como las siguientes:

- 1.- Todas las respuestas registradas en los formularios como **NR**, **NS**, **NI** significan (no responde, no sabe, no informa) y lo que se ingrese será el código **9**, si el campo es de un dígito. Si el campo es de dos dígitos **99**, si el campo es de 3 dígitos **999** y así sucesivamente.

- 2.- Si el campo esta en blanco, **digite +** (en el caso de aplicaciones IMPS).
- 3.- Nunca forzar el ingreso de un valor no permitido.
- 4.- Si en un campo numérico el espacio no es suficiente para la cantidad declarada, digitar nueves y en el último dígito ponga ocho. Ejm.

Si el campo es de 3 dígitos deberá ingresar 998

Si el campo es de 5 dígitos deberá ingresar 99998

Todo esto para que en la fase de procesamiento y/o explotación de la información, a través de un análisis exploratorio de datos, estos casos estén sujetos a exclusión o imputación de valores válidos según el caso.

IV. Integración de Variables: Gasto, Consumo e Ingreso. Desocupados, Subempleados, etc.

Dependiendo de la naturaleza de cada encuesta y del usuario que adquiera la base de datos, todas las variables agregadas y/o calculadas si se integran en la base de datos, tal es así que:

- En la encuesta de Empleo las bases de datos de todos los años siempre tienen integradas las variables de: Factor de expansión, Ingreso de la persona, sectorización de la PEA y condición de actividad, puesto que a través de estas variables se tiene el espectro completo para analizar los diferentes aspectos que se pueden tratar en una encuesta de este tipo. Cuando un usuario adquiere cualquiera de estas bases, estas siempre van acompañadas de su correspondiente diccionario de datos, en el cual se describen los códigos y etiquetas o significados de los mismos
- En la encuesta de condiciones de vida solamente se integra la variable factor de expansión y en casos especiales dependiendo del usuario toda la gama de variables respecto al ingreso y gasto mensualizados del hogar.
- En la encuesta de ingresos y gastos todavía no se ha definido todavía como se va a proceder al respecto.

V. Arquitectura de la Base de Datos

1. Encuesta de Empleo

- Para cada una de las diferentes encuestas de empleo básicamente se tiene un solo archivo de datos, el mismo que está en formato SPSS.
- Para identificar un hogar, en todas las encuestas utilizamos la clave que se compone de las variables CIUDAD+ZONA+SECTOR+VIVIENDA+HOGAR.
- Para identificar un persona en particular se utiliza la clave que se compone de las variables CIUDAD+ZONA+SECTOR+VIVIENDA+HOGAR+PERSONA, de tal manera que cualquiera de estos archivos de datos, un registro representa a un persona.
- En esta encuesta los archivos de datos siempre tienen variables cuyos nombres son los mismos para todas los años, por lo que los usuarios que regularmente trabajan con estos archivos conocen ya muy bien que o cuales son las variables que tienen tal o cual información.

2. Encuesta de Condiciones de Vida

- Tomando como antecedente que las bases de datos originalmente están en formato ASCCI y con un diseño propio de IMPS, necesariamente hay que realizar un proceso intermedio que permite separar los registros de un TIPO específico en una tabla de datos SPSS, de tal

manera que si se tienen tantos tipos de registro, existirán tantas tablas de datos; por lo que en el caso de esta encuesta se determinaron las siguientes tablas de datos SPSS: VIVIENDA, PERSONAS, EQUIPAMIENTO, GASTOS EN ALIMENTOS, OTROS GASTOS, NEGOCIOS, ACTIVIDADES AGROPECUARIAS.

- En cualquiera de las tablas a un hogar se identifica a través de la clave compuesta CIUDAD+ZONA+SECTOR+VIVIENDA+HOGAR.
- Todas las variables excepto las que forman parte de la clave, siguen un estándar a través del cual se puede relacionar a la variable con tal o cual pregunta del formulario. Ejemplo:

VI03 significa sección vivienda pregunta 03

PE04 significa sección personas – educación pregunta 04

PA02 significa sección personas – actividades económicas pregunta 02

3. Encuesta de Ingresos y Gastos

- Esta encuesta por su naturaleza tiene varias tablas de datos, entre las que tenemos son: HOGARES, PERSONAS, PERCEPTORES, GASTOS DIARIOS, GASTOS MENSUALES, GASTOS ANUALES.
- Al igual que en el resto de encuestas siempre se identifican a los hogares a través de la clave compuesta CIUDAD+ZONA+SECTOR+VIVIENDA+HOGAR
- Los nombres de las variables, para una fácil manipulación de las tablas de datos, siguen un estándar similar al que se utilizó en condiciones de vida.

4. Encuesta de Indicadores de la Niñez y los Hogares

- En esta encuesta se procedió a generar tablas de datos SPSS al igual como se realizó en las de condiciones de vida, puesto que en esta también se utilizó el IMPS para el ingreso de datos.
- Así mismo en cuanto a las variables que permiten identificar de manera única a un hogar o persona se procedió como se lo ha hecho para el resto de encuestas.

VI. Generación de Tabulados

Todo lo que es generación de tablas y en muchos casos el cálculo de variables agregadas se realizan en SPSS. Casi en un 100% se generan programas en lenguaje SPSS a fin de ser re-utilizados, situación que permite oportunidad y rapidez en la elaboración de cuadros.

VII. Políticas Institucionales de Acceso a las Bases de Datos

En base al objetivo de cada una de las encuestas de hogares, generalmente las políticas institucionales adoptadas para que los usuarios accedan a las base de datos son:

1. Vender las bases de datos a precios mínimos
2. Asesorar a las personas e instituciones usuarias, en el manejo de las bases.
3. Vía convenios institucionales, hacer conocer a nivel de todo el país la existencia de estas bases.
4. Elaborar publicaciones de carácter descriptivo a fin de promover la demanda de estas bases.

**METODOLOGÍA PARA LA INTEGRACIÓN DE BASES
DE DATOS DE LA ENCUESTA NACIONAL DE
EMPLEO E INGRESOS - ENEI**

INSTITUTO NACIONAL DE ESTADÍSTICA (INE)

GUATEMALA

ÍNDICE

	<u>Página</u>
Presentación	125
Esquema General de Procesamiento de Datos - ENEI.....	125
Capítulo I.....	126
1. Aspectos Metodológicos Aplicados en la Captura de la Información en Campo.....	126
1.1 Etapas de captura en campo	126
Capítulo II.....	129
2. Etapas de Trabajo en Oficina	129
Capítulo III	131
3. Aspectos Relacionados con la Integración de la Base de Datos para la Generación de resultados	131
3.1 Etapa s de generación de indicadores.....	131

Presentación

El Instituto Nacional de Estadística -INE- de Guatemala, es una institución descentralizada, y semiautónoma que tiene por objeto formular y realizar la política estadística nacional así como planificar, dirigir, coordinar y supervisar las actividades del Sistema Estadístico Nacional –SEN; entre sus principales atribuciones, se encuentra la de ejecutar los Censos y Encuestas Nacionales (Ley Orgánica, Decreto Ley 3-85, artículo 28),.

El INE presenta este documento como un aporte a la discusión de la temática programada en el marco del Undécimo Taller Regional de MECOVI.

Cabe mencionar que en el contexto del MECOVI, el Gobierno de Guatemala y el INE, realizaron en el año 2000 la primera encuesta sobre Condiciones de Vida ENCOVI; y en el 2002 y 2003, la primera ronda de cuatro encuestas trimestrales sobre empleo e Ingreso –ENEI-, encuestas que han permitido generar información socioeconómica integral y detallada sobre los niveles de empleo, subempleo, desempleo e ingresos de los hogares y personas, según áreas y etnias del país.

El esfuerzo de realizar ENCOVI y ENEI, (como parte principal del Sistema Nacional Integrado de Encuestas de Hogares), vino a llenar un vacío manifiesto en la producción de información sobre el tema del mercado laboral en Guatemala, al mismo tiempo de posibilitar el estudio de las tendencias, mecanismos y dinámicas relacionadas con los procesos de producción y empleo, analizar las causas y factores que determinan el desempleo y subempleo y establecer los alcances e impactos de los programas de generación de puestos de trabajo y monitorear los avances y logros de las acciones sociales del estado en el campo de la política laboral.

La ENEI está diseñada como respuesta a la necesidad de garantizar la caracterización y análisis integral de las variables relacionadas con los temas de actividad económica e ingresos de la población de Guatemala.

Para su mejor comprensión, este documento está dividido en tres capítulos: el primer capítulo se presenta el esquema relacionado con la captura de información en campo; el segundo capítulo contiene las especificaciones del trabajo de oficina y el último capítulo el proceso de la generación de resultados.

Se espera que este documento sirva de insumo y como medio de compartir la información de Guatemala con los colegas reunidos en el Undécimo Taller sobre Metodologías para la Integración de las Bases de Datos a realizarse en la histórica y bella ciudad de Lima, Perú del 16 al 18 de julio del 2003.

Esquema General de Procesamiento de Datos -ENEI

Con el propósito de sistematizar metodológicamente los procesos técnicos que se realizan en el marco de la Encuesta Nacional de Empleo e Ingresos –ENEI- y de dotar a la institución de una metodología probada y validada que pueda utilizarse en el futuro desarrollo de las encuestas, se consideró conveniente iniciar esta etapa plateándola desde la perspectiva técnica y administrativa del proceso de manejo adecuado de las bases de datos.

Este aporte pretende propiciar canales de comunicación con las demás instituciones de estadística y mostrar la manera como se desarrolla en Guatemala el manejo del procesamiento de datos.

Considerando las características metodológicas de la ENEI se diseñó y puso en operación un programa de procesamiento de datos que consta de varias etapas que van desde la planificación de actividades (dentro de éstas se encuentra la planificación del procesamiento de datos que se inicia desde el momento de los requerimientos de software y hardware y otros insumos necesarios para el desarrollo adecuado del mismo) hasta la producción y publicación de resultados.

La segunda fase, da inicio al momento de contar con el formulario definitivo, a partir del cual se desarrollan los programas respectivos y otros procesos que se describen a continuación:

El proceso del levantamiento de la encuestas se inicia con la recolección de los datos por parte del encuestador. La aplicación de la metodología aplicada en MECOVI, implica la conformación de equipos de campo, constituidos por tres encuestadores, un supervisor, un digitador y el piloto. El supervisor de grupo tiene como tarea central el control de cobertura, calidad, oportunidad, digitación, consistencia y comportamiento de los encuestadores a su cargo. A este equipo se une una figura que se le denomina como Monitor.

Los monitores son los representantes de la Dirección Técnica quienes tendrán a su cargo la supervisión y el monitoreo permanente de los trabajos de la ENEI en el campo. Tienen como función principal ASEGURAR la correcta aplicación de la metodología ENEI; ASEGURAR que las metas semanales de cobertura se cumplan; APOYAR el trabajo técnico de supervisores, encuestadores y digitadores; UNIFICAR la recolección a nivel nacional en términos de procedimientos, técnicas y normas acordadas.

La etapa de ingreso de datos al computador y su posterior verificación se realiza en el campo, y está a cargo de los Digitadores. Estas personas, al igual que los encuestadores y supervisores, son cuidadosamente seleccionados y capacitados, teórica y prácticamente, para el correcto desempeño de sus labores. En adición los digitadores son los responsables de la administración y cuidado del equipo de procesamiento.

Capítulo I

1. Aspectos Metodológicos Aplicados en la Captura de la Información en Campo

1.1 Etapas de captura en campo

a) Codificación

Con el propósito de facilitar el procesamiento de los datos en la boleta han sido codificada la mayoría de variables investigadas; para algunas que incluyen la categoría “otros” necesariamente se debe realizar un proceso de codificación igual situación se presenta para otras preguntas abiertas de los formularios, para las que debe determinar la clase a la que pertenecen y asignar el código que las identifica.

Este proceso se realiza en la mayoría de los casos por el supervisor durante la revisión de los formularios, antes de su entrega al digitador para su procesamiento. El supervisor es el responsable de la codificación de los temas: Departamento y municipio, Rama de Actividad Económica y Ocupación y otros que se consideren pertinentes. Se ha implementado en el sistema la captura de las descripciones texto de las preguntas abiertas para luego hacer una recodificación semi-automática.

b) Captura de datos

La captura de los datos de la encuesta se hace utilizando el paquete Centry de IMPS, que permite capturar la información correspondiente a cada uno de los hogares del sector, y al interior de cada hogar la información específica de cada persona. Para ello al momento de la captura se distribuye la información de un hogar en varios archivos independientes, cada uno de ellos constituirá un Batch y en general cada parte de la información del hogar se constituye en un Formulario.

El programa de captura realiza, en una secuencia de pantallas, un barrido de todos los ítems investigados, en el orden en que aparecen en el formulario de hogares ENEI, respetando los flujos establecidos.

Durante el proceso de captura se realiza una verificación de rangos de todas las preguntas del formulario para establecer que se encuentren dentro de los límites establecidos. Dada la

incorporación del proceso de captura al trabajo de campo, no se efectúa ninguna verificación de doble digitación. Este proceso se estima cambiará en el 2004 por el Cspro del Bureau de Censos.

c) **Consistencia de datos**

La etapa más compleja y laboriosa del procesamiento de datos del proyecto es sin duda la consistencia de los datos, la cual tiene como propósito garantizar la calidad de la información recolectada, mediante la revisión sistemática de cada una de las respuestas tanto de manera individual como en su relación con otras.

Para efectuarla se toman como base las pautas de consistencia definidas por los especialistas sustantivos de cada área temática del proyecto. Las mismas se clasifican como errores o como advertencias; los errores, que reflejan relaciones incoherentes por definición, son objeto de corrección siempre; las advertencias, que reflejan incoherencias aparentes en una relación, son corregidas o aceptadas dependiendo del resultado de la evaluación realizada por el personal de campo.

El proceso de consistencia genera listados de error, los cuales sirven para que el personal de campo pueda efectuar las revisiones y análisis que correspondan, hasta determinar el origen del error, establecer la corrección a efectuar y realizar la actualización de los archivos; también sirven como soporte documental de los procesos de consistencia efectuados, de manera que los monitores y directores del proyecto puedan evaluar en cualquier momento, durante el desarrollo del trabajo, los problemas de calidad que se están presentando, las soluciones que se dieron y tomar con base en su análisis las acciones correctivas y de estandarización de procesos que correspondan.

Por otra parte, estos listados son un valioso material de consulta para los analistas de resultados y especialistas temáticos, interesados en conocer a detalle el grado de cambio que el proceso consistencia de datos introdujo en los resultados.

Al emitir un listado de error, el digitador verifica, mediante un proceso de comparación contra boleta y archivos, que los errores señalados no son producto de una mala digitación, debiendo señalar en el listado, las correcciones que procedan de esta fuente y entregar el listado al supervisor para su revisión; El supervisor revisa el listado contra boletas para establecer la validez de las anotaciones del digitador y validarlas con su visto bueno; luego revisa el resto de los errores si los hay, procediendo a entregarlo al encuestador para que este a su vez revise la información, y según el caso haga las correcciones o vuelva al hogar correspondiente a rectificar la información con problemas. El encuestador al terminar de señalar las correcciones en el listado, lo devuelve al supervisor, para que este nuevamente lo revise y valide las correcciones señaladas, luego lo entrega al digitador para que este realice la corrección de los errores y repita los procesos de consistencia, hasta obtener listados sin errores. (Ver diagrama.)

El proceso de consistencia consta de varios módulos, los cuales se explican a continuación.

d) **Chequeo de cobertura de viviendas**

Con base en los listados de actualización cartográfica de los sectores seleccionados, se preparó el marco muestral del proyecto; a cada encuestador y supervisor se le proporciona un listado de los segmentos seleccionados del sector que contiene la información de todas las estructuras del segmento, con la especificación de la situación de cada una de ellas al momento de la actualización cartográfica, este listado es revisado en terreno por el supervisor de campo y actualizado con las modificaciones, adiciones o eliminaciones que corresponda. Posteriormente, cuando cada segmento ha sido capturado, se efectúa un proceso de pareo que permite validar que todas las viviendas y hogares del segmento han sido capturadas, y que no existen omisiones.

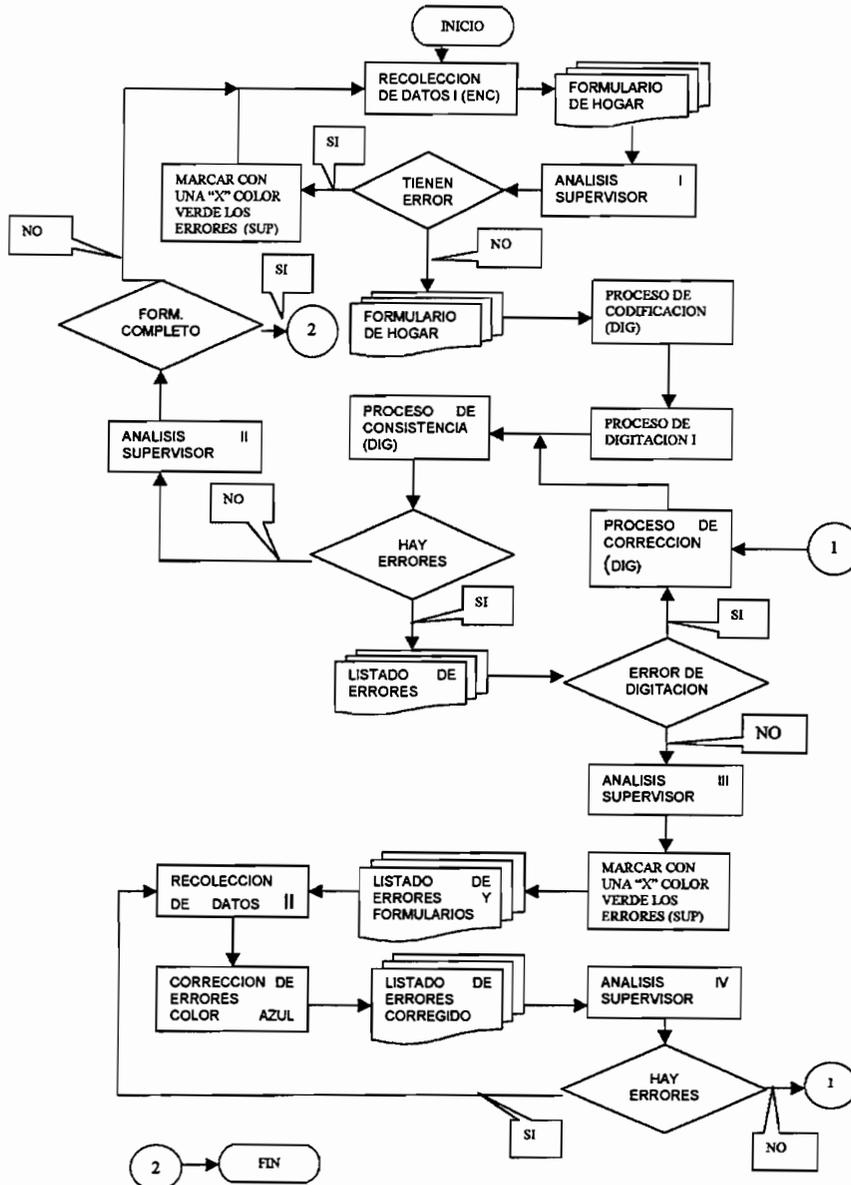
e) **Chequeo de integridad de hogares**

Como se indicó anteriormente, la captura de datos desagrega la información de un hogar en varios archivos, para establecer su integridad; durante el proceso de revisión que efectúa el supervisor de campo, se establecen las cifras de control del hogar, las cuales se registran en un formulario específico y con los datos de éste se construye un archivo, el cual permite efectuar un análisis de integridad (completitud) de cada hogar.

f) **Validación de rangos**

Adicionalmente al proceso de verificación de rangos que se efectuó al momento de la captura de los datos, al tener los archivos grabados, se efectúa un proceso de validación de rangos, que tiene como propósito establecer la validez individual de la respuesta del informante a cada pregunta, eliminando los errores provenientes de la utilización de captura forzada, y los eventuales errores introducidos durante los procesos de corrección.

DIAGRAMA DEL PROCEDIMIENTO DE CONSISTENCIA DE DATOS



Descripción del Procedimiento de Consistencia de Datos

1. El encuestador recoge la información del hogar seleccionado
2. El encuestador entrega al supervisor de grupo los formularios que tienen información pendiente de procesar.
3. El supervisor revisa los formularios entregados por el Encuestador y analiza si hay errores por parte del encuestador, los marca con una "X" de color verde. Si hay errores por parte del encuestador, el supervisor regresa los formularios con error al encuestador para que los corrija directamente en los hogares.
4. El supervisor de grupo entrega los formularios al Digitador para que este los capture en la Lap Top proporcionada para el efecto.
5. El Digitador ejecuta el procedimiento de Consistencia de Datos en la opción de menú que corresponda en el sistema de procesamiento de datos.
6. Si el sistema emite el listado de errores correspondientes, el Digitador verifica si hay errores de "Digitación". Si hay errores de Digitación, el digitador graba en la opción de menú correspondiente, los errores detectados.
7. El supervisor recibe del digitador el listado de errores de inconsistencia y los respectivos formularios; analiza el listado y si el formulario fue procesado sin error va a paso No. 12, si no, marca con una "X" de color verde, los errores a corregir.
8. El Encuestador recibe los formularios y el listado de errores para visitar de nuevo a los hogares y corregirlos. Si la corrección es un cambio en un valor, el mismo se hace a la par (en el formulario y el listado) con color azul, si no hay ningún cambio, es decir permanece el mismo valor, el encuestador hace un cheque a la par del dato (en el formulario y el listado) correspondiente con color azul.
9. El supervisor recibe de nuevo el listado de error y formulario. Revisa que todos los errores hayan sido corregidos y que la corrección se haya anotado tanto en el formulario como en el listado.
10. El supervisor entrega al digitador el listado de errores corregido y los formularios que correspondan.
11. El digitador captura las correcciones señaladas, las aplica a los archivos correspondientes y realiza de nuevo el procedimiento a partir del paso numero 5.
12. Si el formulario fue procesado completamente, el supervisor notifica verbalmente de esta circunstancia al digitador y al encuestador correspondiente y almacena el formulario en el lugar que corresponda; Si el formulario no esta procesado completamente, debe regresarlo al encuestador para que se repita este procedimiento a partir del paso No. 1 y completar la información pendiente.

Capítulo II

2. Etapas de Trabajo en Oficina

a) Integración de archivos

La integración de la captura y consistencia básica de los datos al proceso de campo, implica una captura descentralizada de los datos, lo que hace necesario un proceso de integración de archivos en dos etapas, la primera al interior de cada sector, que integra los datos de los hogares (desagregados para la captura), en un único archivo del sector y la segunda que integra todos los sectores trabajados durante la comisión en un único archivo del proyecto. Para ello se ejecuta en la primera etapa, un proceso que concatena y reordena los datos al interior del sector y dentro de estos al interior del

hogar y en la segunda etapa un proceso que une y reordena los datos al interior del país y dentro de este al interior de cada dominio de estudio.

b) Reestructuración de los datos

Los datos se capturan respetando el orden en que aparecen en el formulario y la secuencia de recolección y con el propósito de facilitar su captura, cada formulario se desagrega en distintos archivos; para ejecutar con los datos los procesos de consistencia y actualización; se realiza una primera reestructuración de los datos, que convierte los archivos planos en bases de datos, y los organiza para facilitar su almacenamiento, recuperación y proceso.

c) Generación de cifras de control

Durante el procesamiento se produce un conjunto de resultados parciales que permite realizar el seguimiento de los procesos de integración y reestructuración de archivos, de manera que se puede garantizar la no ocurrencia de omisiones y/o duplicaciones derivadas de procesos mal aplicados. Para ello se producen listados de distribución de frecuencias simples, antes y después de los procesos señalados, para las variables que se definen como básicas para el seguimiento de los procesos.

d) Relaciones entre variables

La validación de la coherencia de una respuesta, valida individualmente, con otra también igualmente válida de manera individual, es sin duda la tarea más compleja y delicada del proceso de consistencia de los datos de ENEI; para efectos de facilitar la comprensión de esta fase de la consistencia que se da entre las variables, se agrupan de la siguiente manera:

▪ **Control de flujos**

La revisión de flujos establece la integridad temática al interior de cada hogar, persona, revisa que se cumplan los universos (sexo y edad) y pases (condicionales e incondicionales) que se especifican en el formulario de hogares. Para ello se verifica que no exista información en las preguntas que no son aplicables en función del universo.

▪ **Cuadre de totales**

Una revisión importante es la que se refiere a la consistencia de los totales de los distintos ítem de valores del formulario; se verifica en algunos casos que los valores parciales estén comprendidos en el total y en otros que la suma de los parciales cuadre con el total.

▪ **Relación lineal**

La revisión lineal establece la coherencia básica de las respuestas individuales, dentro de un hogar, persona. Para ello se verifica el cumplimiento de las pautas de relación lineal (simple) de una respuesta con otra u otras respuestas, normalmente de una misma área temática.

▪ **Relación estructural**

La revisión de estructural establece la coherencia global de las respuestas individuales al interior del hogar en su conjunto. Relacionando las respuestas de un área temática con los de otra u otras áreas temáticas. Para ello se verifica el cumplimiento de las pautas de relación estructural (compleja) de una respuesta con otra u otras respuestas, normalmente de distintas áreas temáticas.

Posteriormente y en la medida en que se completa el proceso de integración, los datos son objeto de otro proceso de reestructuración, que comprende entre otras cosas, la construcción de variables derivadas, la incorporación del factor de expansión y la preparación de las bases de datos para análisis de resultados, por parte de los especialistas sustantivos, en ambiente SPSS.

Capítulo III

3. Aspectos Relacionados con la Integración de la Base de Datos para la Generación de Resultados

3.1 Etapa de generación de indicadores

La ENEI desarrolla un amplio conjunto de indicadores del mercado laboral a partir de la base de datos de cada encuesta como principal fuente de información estadística del país, elaborados mediante métodos comparables internacionalmente y sujetos a un cuidadoso control de calidad.

El proceso de cálculo es sistemático: empieza con la utilización de las definiciones y conceptos de la Organización Internacional del Trabajo -OIT- y concluye con la presentación de los indicadores del mercado laboral en un formato de fácil interpretación para los usuarios en general.

Para elaborar los indicadores del mercado laboral se siguen los siguientes pasos:

a) **Definición.**

El primer paso en la elaboración de un indicador del mercado laboral es la consideración de su definición conceptual (OIT) como una aproximación o una medida de un fenómeno o variable, así como la aplicación de criterios nacionales. Para obtener una medida correcta se delimita de manera precisa el fenómeno que interesa estimar.

b) **Fuentes de información**

La fuente primaria de investigación la constituye la ENEI misma y como fuente secundaria de comparación las ENEI anteriores u otra encuesta anterior, así como los censos de población y registros administrativos que contienen información relacionada con esta materia. En ocasiones se utilizan también fuentes provenientes del Acuerdos Ministeriales y de otras instituciones del sector público y privado.

c) **Población de referencia.**

Es importante indicar que todo indicador debe referirse a una determinada población que se define según su propósito de medición; para la ENEI, se obtiene información de personas de 7 años y más de edad.

La mayoría de los indicadores del mercado laboral tienen como base a la población en edad de trabajar (PET) del país; otros indicadores como los de educación y empleo, consideran a grupos específicos (por ejemplo, grupos de edad u ocupación, rama de actividad, categoría ocupacional, etc.).

d) **Tabulación básica**

Aún cuando la naturaleza del procesamiento no se enfoca a la obtención tradicional de plan de tabulaciones, se realiza un proceso de tabulación para un conjunto de cruces considerados básicos para la evaluación y difusión de resultados. Este proceso se efectúa con valores muestrales y expandidos.

e) **Limpieza de la base de datos primaria**

Se denomina bases de datos primarias al conjunto de datos contenidos en un formato generado posteriormente al proceso de digitación de los datos en el campo y su respectiva consistencia.

Esta base de datos preliminar por ser generada inicialmente con propósitos de utilizarla como insumo básico para el cálculo de los resultados finales de la ENEI es objeto de un tratamiento especial y de riguroso cuidado que se inicia por verificar y cuadrar universos de variables por capítulos, así como estructuralmente la completitud de la base de datos.

Se presume que la base de datos preliminar puede contener errores de digitación y de codificación así como inconsistencias, por lo que se efectúan procesos de verificación y limpieza exhaustiva de las variables utilizadas, aspecto que incluye: el examen de la distribución de cada variable, la generación de listados y su verificación en boleta así como la corrección de la misma; también se observan los errores de digitación y el tratamiento de casos atípicos denominados (“outliers”) y de valores faltantes (falta de respuesta) o información no aplicable y casos agrupados en clasificaciones como ¿otro cual?, etc, para la obtención de la limpieza de la base de datos final.

f) Generación de los expansores para la base de datos final

Para el cálculo de los factores de expansión se aplican los recíprocos de las probabilidades de selección en la muestra (fracciones de muestreo). Este proceso se vincula con la base de datos final a la cual se incorporan los expansores respectivos.

g) Ingreso a la base de datos final

El indicador resultante y los expansores correspondientes se ingresan en una base de datos terminal en su máximo nivel de desagregación. Esta base de datos es a su vez, manejada por el paquete informático Statistical Package for the Social Sciences (SPSS).

h) Generación de indicadores

Con la base de datos final se procede a programar las fórmulas de cálculo de los indicadores del mercado laboral y agregados de ingreso o consumo según se requiera. Para estos procesos se utiliza el paquete SPSS.

i) Control de calidad de los indicadores

Consiste en la revisión cuidadosa del programa de cálculo y los resultados. En esta etapa se verifican:: a) la creación de tablas y cruces específicos de variables por ocupación, rama de actividad, categoría ocupacional, grupos de ocupación, educación, etc.) b) los resultados del cálculo de los indicadores según las distintas fuentes existentes y con resultados publicados c) el indicador para distintos grupos de población con diferentes desagregaciones.

j) Control de calidad final de la base de datos

El INE valida los indicadores provenientes de información de sus bases de datos. Este proceso termina con la revisión de la fidelidad de la base de datos, corroborando en pantalla de salida los indicadores del mercado laboral utilizando SPSS.

k) Control de calidad final de la presentación de los resultados

Esta fase considerada como la culminación del dato que se presentará a los usuarios y por representar la imagen del proyecto, requiere de una revisión cuidadosa previa a emitirse el visto bueno para la generación de la base y la publicación respectiva.

l) Presentación y difusión de resultados

Con el propósito de que los resultados de la encuesta puedan ser de fácil difusión, recuperación y consulta, se realiza un proceso de preparación de bases de datos que permitan y viabilicen la recuperación y consulta de los resultados de la encuesta en sus diferentes áreas temáticas.

De igual forma se prepararan los productos que se definan como objeto de difusión, utilizando las vías de internet: www.ine.gob.gt, así como Intranet y la disponibilidad de bases y resultados para los diferentes usuarios.

**ENCUESTA NACIONAL DE EMPLEO E
INGRESOS – ENEI 2002**

INSTITUTO NACIONAL DE ESTADÍSTICA (INE)

GUATEMALA

ÍNDICE

	<u>Página</u>
Primera Etapa: Trabajo de Campo	135
I. Estrategias para la captura de datos	135
II. Procesamiento para evaluar la consistencia de la información.....	135
III. Diseños computacionales de las bases de datos	136
Segunda Etapa: Trabajo en las Oficinas Centrales.....	136
IV. Recibimiento de la información e integración de los archivos.....	136
V. Segunda etapa de consistencia: información consistente y coherente.....	136
VI. Imputación de datos	137
VII. Integración de variables	138
VIII. Arquitectura de las bases de datos.....	138
IX. Generación de tabulados	138
Tercera Etapa: Difusión de Resultados y Acceso a Bases de Datos	139
X. Políticas institucionales de acceso a bases de datos	139

Primera Etapa: Trabajo de Campo

I. Estrategias para la captura de datos:

- **¿Se realiza en oficinas o de forma descentralizada?**

La captura de datos es un proceso realizado directamente en el campo, haciendo uso de computadores portátiles. Cada equipo de trabajo (12 en el caso de la ENEI) cuenta con un supervisor de grupo, tres encuestadores y un digitador.

- **¿Cómo se codifican las preguntas?**

La boleta de la encuesta está previamente codificada, lo que permite una fácil captura de la información.

Prácticamente toda la información es ingresada de manera codificada desde un inicio, a excepción de algunas variables como: rama de actividad económica, título o diploma obtenido, tipo de cursos de capacitación. Éstos son codificados en las oficinas centrales, según clasificaciones internacionales (CIU,...).

II. Procesamientos para evaluar la consistencia de la información:

Los grupos de trabajo están compuestos por supervisores de grupo, encuestadores y digitadores; respaldados y supervisados por los supervisores de procesamiento y monitores. Las obligaciones de éstos son:

Monitores: son los representantes de la Dirección Técnica y deben velar por la calidad de la información recabada. Sus funciones principales son (auditoria de campo):

- Asegurar la correcta aplicación de la metodología de la encuesta.
- Asegurar que las metas semanales de cobertura sean cumplidas.
- Apoyar el trabajo técnico de los supervisores, encuestadores y digitadores.
- Elaborar informes semanales sobre el estado de la encuesta con el fin de implementar los apoyos y correctivos donde sean necesarios.

Supervisores de procesamiento: Tienen como tarea principal, asegurar el funcionamiento del proceso de captura de datos, efectuar las actualizaciones de los sistemas que se requieran y dar soporte técnico a los digitadores.

Supervisores de grupo: El supervisor de grupo tiene como tarea central el control de la cobertura, calidad, oportunidad, consistencia y cumplimiento de los encuestadores a su cargo.

Encuestadores: La recolección de los datos estará a cargo de los encuestadores, los que serán cuidadosamente seleccionados y capacitados, tanto en teoría como en práctica.

Digitadores: Cada grupo de trabajo cuenta con un digitador, quien tiene a su cargo el ingreso de la información recolectada por los encuestadores.

- **¿Se validan los datos cuando se capturan?**

Luego de una primera revisión de la boleta por el supervisor, la información es ingresada en el campo a medios electrónicos usando un programa de captura en IMPS. Para una primera etapa de validación y consistencia de la información, se utiliza un programa en FOX PRO el cual, por medio de pautas, permite la identificación de errores. Las pruebas de consistencia se realizan de manera periódica y sistemática a lo largo del proceso de captura, se generan reportes y se procede

a detectar si el error fue de digitación o de incorrecta recolección, si fuera un error de recolección se debe regresar al hogar y entrevistar nuevamente a la persona(s).

III. Diseños computacionales de las bases de datos:

- **¿Qué software se utiliza para la captura, validación, procesamiento y explotación de las bases de datos?**

Captura de datos: IMPS

Validación en campo: FOX PRO

Procesamiento, revisión de consistencia y explotación de las bases: SPSS

- **¿Existe software de desarrollo propio?**

Tanto el IMPS como el FOX PRO debieron adecuarse y ajustarse para los objetivos y parámetros específicos de la encuesta.

- **¿Cómo se justifica el uso de los sistemas vigentes frente a otras opciones existentes en el mercado?**

Una de las principales razones de la utilización de estos sistemas, es que fueron éstos los empleados en la Encuesta Nacional de Condiciones de Vida, realizada en el año 2000, lo que facilitaba el proceso de implementación de la ENEI.

El uso de otros sistemas no ha sido descartado, en el trabajo de campo de la última ENEI realizada, el CS-PRO fue probado paralelamente al sistema tradicional.

Segunda Etapa: Trabajo en las Oficinas Centrales

IV. Recibimiento de la información e integración de los archivos:

Finalizado el trabajo de campo, se recibe los archivos en las oficinas centrales y se procede a la integración de los mismos. Luego de algunos procesos de consistencia en las bases de datos integradas, se realiza la importación de datos a formato .sav (SPSS), conformando 2 archivos: uno de hogares y otro de personas.

V. Segunda etapa de consistencia: información consistente y coherente:

La segunda etapa de consistencia de la información se realiza en las oficinas del Ine, utilizando las bases de datos y el programa SPSS. Este procedimiento se realiza para cada una de las variables y consiste en:

- Revisión de los universos: Que los casos válidos sean los correctos y que se respeten los flujos de la boleta.
 - Ej. Que todos (y solo) las personas clasificadas como trabajadores asalariados (empleados del gobierno, trabajadores del sector privado, empleadas domésticas y jornaleros) tengan información en las variables “sueldo o salario mensual”, “prestaciones laborales”, etc. (respeto de flujos).
- Cruces de variables: Elaboración de listados de los casos con información “incoherente” para que sean revisados contra boleta.
 - Ej: Personas con relación de parentesco con el jefe de hogar “cónyuge” y edad menor a 15 años.
 - Ej: Jornaleros o trabajadores domésticos con ingresos muy altos.

Esta revisión se realiza para detectar y corregir los errores de digitación. Para los casos confirmados como errores al momento de digitar, se procede a corregir la base de datos, no así para aquellos donde la información viene en boleta de esta manera.

- Codificación de las variables texto y tratamiento de la categoría “otros”: En este paso se revisan y codifican las variables de texto para agregarlas codificadas a la base de datos. Si para algunas variables (preguntas cerradas con opciones de respuesta) la categoría “otros, cuál?” representa más del 5% se procede a “abrir” la información, creando así nuevas categorías

VI. Imputación de datos:

• Como se trata la no respuesta total o parcial?

La no respuesta total (personas que definitivamente no dieron información o no quisieron concluir la encuesta) es excluida de la base de datos, pero es contabilizada para el cálculo de la tasa de rechazo.

La no respuesta parcial, como la ocurrida en las variables de ingreso, es identificada con un código especial (99999), para luego imputarle un valor. Es importante recalcar que nunca se realizan imputaciones en las variables originales, para esto se crean nuevas variables identificadas con el sufijo “i” (de “imputadas”). Por ejemplo: p05a14 (sueldo o salario) y p05a14i (sueldo o salario con imputaciones).

La imputación de datos es mínima, se trata de mantener la información lo más apegada a la dada por el informante.

• Cuales son los criterios y procedimientos para la imputación de las variables?

Para cada una de las variables que componen el agregado de ingresos, se lleva a cabo el siguiente procedimiento:

Detección de los casos raros

Por medio de una frecuencia simple se detectan valores ignorados (por desconocimiento o rechazo del informante) y valores fuera de rango.

Determinación de los casos extremos

Para determinar los valores extremos se realizaron las siguientes pruebas:

- a) Utilizando una de las herramientas del SPSS, se efectúan test de normalidad para analizar la distribución de los datos, con el objetivo de detectar valores extremos.
- b) Se construye un indicador¹ para determinar a cuántas desviaciones estándar de la media se sitúa cada uno de los valores. Para algunas variables, como por ejemplo “sueldo y salario”, se realizan medias por grupos relativamente similares (se obtienen medias según la categoría ocupacional y el dominio de estudio). De manera que se analicen los datos según los ingresos de las personas con un perfil similar.

Por medio de estas dos herramientas se detectan los valores anómalos y se analizan detenidamente, haciendo cruces de variables para entender si se traga se trataba de un valor extremo o de un valor atípico pero racional.

¹ Indicador : el valor absoluto de : $\frac{\text{Valor} - \text{Media}}{\text{Desviación estándar}}$

Cálculo de los promedios a imputar

Luego de detectados e identificados los valores extremos, se calculan promedios para imputarle a estos y a los casos raros (ignorados, fuera de rango). El promedio a imputar excluye los casos extremos y se efectúa para subgrupos con similitudes, con el objetivo de que el valor imputado se apegara lo más posible a un valor real, según las características de la persona.

Imputación

Finalmente se procede a imputar los casos en una variable nueva que lleva el sufijo “i” (de imputada), a manera que la variable original no sufra alteraciones.

VII. Integración de variables:

• **Se integran las variables agregadas?**

Las principales variables agregadas (en el caso de los ingresos) y construidas (en el caso de las variables de empleo (ocupados, desocupados, subempleados) son integradas a la base de datos, para facilitar el análisis y la investigación que puedan realizar los usuarios de las bases.

• **Cuales son los mecanismos para su normalización en el tiempo?**

Los ingreso se unifican al periodo de referencia de la principales variables (sueldo y ganancias mensuales), en el caso de la ENEI, las variables que compone el agregado de ingreso y que fueron captadas para un periodo de referencia mayor (aguinaldo, remesas trimestrales...) son mensualizadas.

• **Se documenta claramente los procesos de integración?**

El “label” o “etiqueta” de cada variable integrada, especifica su composición.

VIII. Arquitectura de las bases de datos:

• **Como se integran los archivos de datos?**

Para la ENEI se integran dos archivos de datos: de personas y de hogares.

• **Cuales son las estructuras utilizadas para facilitar su manejo en paquetes estadísticos?**

Se presentan como archivos rectangulares para ser manejados con el SPSS. Las bases de datos reflejan la misma estructura y orden de la boleta.

Para ayudar a los usuarios de las bases de datos, Mecovi Guatemala ha impartido cursos sobre manejo y análisis de las bases de datos con SPSS. En noviembre pasado, se dictó el curso: “Introducción al análisis de las bases de datos con el paquete SPSS”, al cual asistieron investigadores del sector público, de centros de investigación y universidades del país.

• **Como se documentan los procesos?**

Las bases de datos se acompañan de Metarchivos, donde se documentan los procesos de la encuesta, la información recolectada (boleta), y la estructura de la base de datos.

IX. Generación de tabulados:

- La generación de tabulados con los principales resultados de la encuesta, se realiza utilizando la herramienta “tables” del SPSS.

Tercera Etapa: Difusión de Resultados y Acceso a Bases de Datos

- Debido a las dificultades para publicar en el sitio web del Instituto Nacional de Estadística los resultados de la encuesta, se realizan envíos masivos (presentación en Power Point y documento en Word) a los usuarios de la información: centros académicos, centros de investigación, medios de prensa, cooperación internacional, ministerios e instituciones del sector público, gremiales e instituciones del sector privado.

X. Políticas institucionales de acceso a bases de datos:

- Difusión de resultados:

Debido a las dificultades para publicar en el sitio web del Instituto Nacional de Estadística los resultados de la encuesta, se realizan envíos masivos, vía e-mail, a los usuarios de la información: centros académicos, centros de investigación, medios de prensa, cooperación internacional, ministerios e instituciones del sector público, gremiales e instituciones del sector privado.

- Acceso a las bases de datos:

La venta o distribución de las bases de datos no es una competencia del Programa MECOVI, por lo que queda a criterio del INE.

Generalmente las bases son vendidas a un precio relativamente bajo, pero no existe una política definida de acceso a las bases de datos.

Contactos en Mecovi en Guatemala:**Director Técnico:**

Carlos Mancía Chúa

cmanciachua@yahoo.com

Analistas:

Erick Chuquiej

chuquiej01@hotmail.com

Mauricio Guerra

mauricio23760@hotmail.com

Vivian Guzmán

guzmanvivian@yahoo.com

Informáticos:

César Calderón

calderces@hotmail.com

Armando Contreras

armandodavidcontreras@yahoo.com



**PROCESAMIENTO DE DATOS DE LA ENCUESTA
PERMANENTE DE HOGARES DE
PROPÓSITOS MÚLTIPLES**

INSTITUTO NACIONAL DE ESTADÍSTICA (INE)

HONDURAS

ÍNDICE

	<u>Página</u>
Introducción.....	143
Aspectos Generales	143
Procesamiento de los datos.....	143
Captura de datos	144
Codificación	144
Digitación	144
Consistencia de la información	144
Diseño de la base de datos.....	145
Validación de los datos.....	145
Integración de las variables	146
Arquitectura de las bases de datos.....	146
Generación de cuadros de salida	146
Programación de tabulados.....	146
Acceso a base de datos	146

Introducción

El Instituto Nacional de Estadística (INE) realiza dos veces por año el levantamiento, procesamiento y presentación de los resultados de la Encuesta de Hogares de Propósitos Múltiples (EPHPM), con el propósito de dar seguimiento a la evolución de las variables socioeconómicas que afectan a la población hondureña.

Dichas variables permiten lo siguiente:

1. Identificar las características sociales y demográficas de la población.
2. Conocer la migración interna.
3. Investigar el número y características de las viviendas en el área urbana y rural.
4. Determinar las tasas de ocupación, subempleo y desocupación de la población.
5. Investigar los ingresos de los hogares generados por trabajo y los provenientes de otras fuentes.

Para la realización de la EPHPM se efectúan las actividades detalladas a continuación:

1. Definición de los objetivos de la encuesta.
2. Revisión del marco conceptual.
3. Elaboración de los instrumentos de recolección de datos.
4. Diseño de la muestra.
5. Capacitación.
6. Realización de la prueba piloto.
7. Levantamiento de datos.
8. Codificación y digitación de los datos.
9. Limpieza de datos.
10. Revisión de los resultados.
11. Publicación

Este informe enfoca solamente los aspectos relacionados con la etapa de procesamiento de datos, la captura, limpieza y publicación de estos, así como las herramientas necesarias para su ejecución.

Aspectos Generales

Procesamiento de los datos

El procesamiento de datos de la encuesta comprende las siguientes actividades:

1. Captura de datos
2. Programación de la entrada de datos y tabulados
3. Codificación
4. Digitación
5. Validación
6. Depuración
7. Generación de cuadros de salida

Captura de datos

La captura de datos es centralizada, ingresada mediante un programa inteligente que aplica criterios de consistencia y validación en línea.

El programa posee pantallas de entrada, similares a las boletas de campo, facilitando el ingreso de los datos. Además, posee un módulo especial que impide la introducción de datos en variables equivocadamente.

La codificación de las variables es numérica, agrupada en tablas correspondientes a cada pregunta de la boleta, lo cual permite su fácil identificación, extracción y relación.

El ingreso de la información se realiza por dominios, es decir, se agrupa por área urbana: principales ciudades, - Tegucigalpa y San Pedro Sula-, resto urbano y el área rural.

Durante el proceso de captura tiene lugar paralelamente el ingreso de los datos para el control de cobertura y poder generar los primeros resultados, a medida que se cubre la totalidad de los dominios.

El programa permite además un control cuantitativo de los errores o inconsistencias mas frecuentes por variable o en relación con ellas.

Codificación

La codificación se inicia una semana después del levantamiento, realizando la identificación numérica de las variables, tales como ocupación, rama de ocupación, países y departamentos territoriales.

Hasta este momento no se realiza corrección alguna, pues es necesario, al momento de generar la información inicial, verificar el trabajo en el campo y comprobar los procesos de capacitación y levantamiento. En todo caso, el programa permite por su parte identificar los errores y reportarlos a los supervisores para su oportuna corrección.

Cada día el supervisor se encarga de revisar las boletas y verificar posibles inconsistencias para corregirlas en el momento; sin embargo, si aun con la revisión quedan errores sin identificar, debido a que las boletas se envían semanalmente, a través del programa de entrada de datos y de las frecuencias generadas en SPSS, ellos pueden identificarse y ser posible corregidos en el campo.

Igualmente se identifica el área geográfica para verificar los segmentos de levantamiento del cuestionario

Digitación

Una semana después de iniciada la codificación, las boletas son ingresadas por el departamento de digitación a razón de 140 boletas diarias por persona, midiéndose la producción y el avance procesado por dominio.

Consistencia de la información

Al momento del ingreso de la información, el programa se encarga de validar los saltos y rangos de valores por variables; el primero se realiza por programa y el segundo desde la base de datos.

Una vez ingresados los datos en el programa de entrada, y luego de enmendar los errores provenientes de la digitación, las base de datos se convierte y se remite al programa SPSS, donde se crean nuevas variables para el procesamiento, como ser: ingreso, condición de actividad, ocupación a un dígito y otras necesarias para la revisión de la información y generación de índices.

Luego se generan frecuencias de las variables de mayor relevancia para observar valores extremos, y posteriormente realizar tablas cruzadas y medir la consistencia de la información.

Este proceso es continuo y próximo a completar cada dominio.

Diseños de la base de datos

La Institución utiliza para la entrada de datos la herramienta Visual FoxPro 6 , para el procesamiento SPSS y para la publicación Microsoft Excell.

El software de entrada de datos es creado por el departamento de tecnología y esta conformado por 6 módulos:

1. Gestión de datos

En este módulo es donde tiene lugar la entrada de datos de la boleta, así como los archivos de control de cobertura, archivos maestros de muestra y todo lo relacionado con la digitación del cuestionario.

2. Procesos administrativos

Donde se lleva el control de acceso a los usuarios, producciones y otros tipos de procesos.

3. Depuración de datos

El programa tienen un mecanismo de validación almacenada en archivos de datos, lo cual permite hacer cualquier tipo de validación en línea.

4. Mantenimiento de parámetros

El programa de entrada de datos necesita de ciertos parámetros que deben proveer las diferentes gerencias.

5. Utilitarios

En este módulo se efectúa el manejo de diferentes opciones en caso de que ocurra un problema con el programa, por ejemplo:

- Respaldo de Información
- Reconstrucción de índices
- Eliminar información de los archivos, etc.

Actualmente, INE está iniciando su proceso de reingeniería a fin de actualizar sus herramientas de desarrollo, planeando utilizar bases de datos relacionales y programaciones de 4ta. Generación para la entrada y divulgación de la información, obteniendo las ventajas que estas herramientas proveen.

Validación de datos

Los datos son validados según matrices de comportamiento de variables, las imputaciones son efectuadas en variables creadas, pues el grado de imputación en la encuesta es bajo y solo se realiza para evaluar resultados del comportamiento de algunas variables.

Si la no respuesta no es significativa, esta se incluye en la presentación de los datos o se descarta la muestra.

Para cada proceso de validación de datos, así como para la creación de variables, se utiliza para su desarrollo SPSS, documentando cada programa.

Los programas continúan creación de variables, ajustes e imputaciones.

Todo cambio efectuado en la base de datos es respaldada, dejando la base original sin alteraciones.

Integración de variables

Todas las variables nuevas se crean de acuerdo a criterios ya sea establecidos a nivel nacional e internacional.

Variables creadas e indicadores son incluidos en la base de datos al momento de su entrega a fin de facilitar al usuario su manejo.

Arquitectura de las bases de datos

La base de datos la conforman las siguientes tablas:

1. Hogares
2. Vivienda
3. Personas
4. Migración

Con la base de datos unificada se procede a la migración del archivo al formato SAV de SPSS llamado Hogar.SAV, dependiendo del año y el mes en que se realizo. En este proceso nuevamente se verifica la consistencia de los archivos.

Generación de cuadros de salida

En el transcurso de la encuesta se generan cuadros de salida por dominio para observar las cifras preliminares y corregir el número de inconsistencias.

Con esto se busca tener el menor numero de errores posible, para el momento de elaborar los cuadros de la publicación.

Programación de tabulados

Para la generación de tabulados es necesario que estén integradas las variables nuevas descritas anteriormente y con esto se procede a su generación en SPSS y luego migrándolos a Microsoft Excel con dos propósitos:

1. Validación cuantitativa de los datos
2. Presentación final de resultados.

Para facilitar el proceso de generación de tabulados y mejorar los tiempos, se han definido macros con Visual Basic para algunos temas de la publicación; otros se realizan a través de vínculos a tablas generadas.

Acceso a base de datos

Las estrategias para el acceso a las bases de las Encuestas de Hogares de Propósitos Múltiples lo puede hacer a través de la Gerencia del Centro de Negocios o por la pagina WEB del INE cuya dirección es la siguiente: <http://www.ine-hn.org> y para cualquier consulta al correo electrónico info@ine-hn.org, inehn@yahoo.com, inehn@hotmail.com.

Las bases se encuentran en SPSS desde el año de 1990 hasta la ultima encuesta publicada el año 2002. También se puede tener acceso a la creación de todo tipo de cuadros, frecuencias, etc., por medio de un software creado con Redatam G-4 Del XVI CENSO de Población y V de Vivienda del 2001.

**DATA WAREHOUSE PARA LA PRESTACIÓN DEL
SERVICIO PÚBLICO DE INFORMACIÓN
ESTADÍSTICA**

RICARDO LUJÁN SALAZAR

INSTITUTO NACIONAL DE ESTADÍSTICA, GEOGRAFÍA E INFORMÁTICA (INEGI)

MÉXICO

ÍNDICE

	<u>Página</u>
Resumen	149
1. Planteamiento.....	149
2. Acerca del Proyecto	150
Objetivos	150
Investigación de proyectos similares.....	150
En qué consiste el proyecto.....	150
3. Arquitectura de Información.....	151
Arquitectura de datos.....	151
Arquitectura de aplicaciones	152
Arquitectura técnica	152
4. Contenido.....	152
5. Desarrollo del Proyecto	153
Extracción, transformación y carga.....	153
Almacén de datos (data warehouse).....	153
Mercado de datos (data marts)	153
Herramientas de acceso.....	153
Metadatos	155
Conclusiones	155

Resumen

El Instituto Nacional de Estadística, Geografía e Informática (INEGI) tiene la responsabilidad de coordinar los Sistemas Nacionales Estadístico y de Información Geográfica de México, además de promover y orientar el desarrollo informático en el país.

Este proyecto consiste en aplicar las tecnologías de bases de datos y *data warehousing* en el desarrollo de un almacén integrado de datos definitivos con información estadística obtenida de los programas de censos nacionales, encuestas y registros administrativos para la elaboración de productos, la toma de decisiones y la planeación facilitando que el personal del Instituto pueda atender con mayor oportunidad los requerimientos de información de los usuarios del INEGI en el marco del Sistema Nacional de Información Estadística y Geográfica.

Palabras Claves: Información Estadística, censos, encuestas, bases de datos, *data warehouse*.

1. Planteamiento

El INEGI, como responsable de coordinar los Sistemas Nacionales Estadístico y de Información Geográfica de México y de promover y orientar el desarrollo informático en el país ha incrementado de forma sustantiva la capacidad de obtener y difundir información mediante la colaboración de personal preparado profesionalmente, el uso de tecnología y su presencia en las 32 entidades federativas de la República Mexicana.

Actualmente el INEGI cuenta con información proveniente de 56 distintos proyectos estadísticos, de los cuales se tiene diferente control y flujo de información, lo que provoca las siguientes áreas de oportunidad:

Concepto	Situación actual	Área de oportunidad
Ubicación de la información.	Islas de información.	Única fuente de datos definitivos con acceso distribuido mediante aplicaciones independientes de la base de datos.
Conocimiento de la información.	Disperso en las áreas generadoras.	Interpretación de la información asistida por metadatos y documentación suficiente.
Congruencia de datos.	Aparente incongruencia entre datos de distintas fuentes.	Obtener interpretaciones de los datos considerando las características particulares de cada evento.
Generación de productos.	Información "amarrada" con las aplicaciones.	Información independiente de las aplicaciones.
Aprovechamiento de la información.	Parcial.	Total.

Dada esta situación se determinó la necesidad de aprovechar las tecnologías de *data warehousing* (extracción, transformación y carga; bases de datos relacionales; bases de datos multidimensionales; metadatos) para mejorar los procesos de generación y explotación de información estadística de modo que la información esté disponible en línea de una manera ágil y entendible para quien lo requiera con el objetivo de apoyar el servicio público de información estadística en beneficio de los diferentes sectores de la sociedad mexicana.

En el segundo trimestre del año 2002 se inició con el desarrollo de la primera etapa del proyecto que comprendió el análisis, diseño y desarrollo de la base de datos para nueve proyectos estadísticos.

2. Acerca del Proyecto

Objetivos

- Mejorar la oportunidad en la prestación del servicio público de información estadística.
- Incrementar la confiabilidad de la información generada por el INEGI.
- Lograr la satisfacción del cliente desarrollando nuevos y mejores servicios de información en línea.
- Facilitar el proceso de toma de decisiones basado en información estadística.
- Reducir los costos de operación en el aprovechamiento de información.

Investigación de proyectos similares

Al inicio del proyecto y durante su desarrollo se ha mantenido un proceso de investigación permanente para conocer y analizar proyectos de *data warehouse* en diferentes instituciones entre las que están:

- Nacional Financiera
- Banco de México
- Comisión Federal de Electricidad
- Buró de Censos de los EEUU

De esta forma se ha podido analizar las prácticas que han dado mejor resultado en estas instituciones para aplicarlas en el proyecto.

En qué consiste el proyecto

Consiste en el análisis, diseño, desarrollo, implementación, mantenimiento y mejora continua del repositorio de información estadística del país para consulta y análisis que permita a los usuarios tomar mejores decisiones como parte del Sistema Nacional de Información Estadística y Geográfica.

Tiene tres elementos principales:

1. Procesos automáticos de extracción, transformación y carga para mantener la información actualizada.
2. Información (datos y metadatos)
 - Integrada
 - Valuada
 - Documentada
 - Única
3. Facilidades de acceso a la base de datos desde cualquier punto de la red INEGI para la consulta, explotación y análisis de la información.

El proyecto no se considera como un producto sino como un procedimiento para integrar y homogeneizar la información proveniente de los proyectos de generación de información estadística, por lo que una

parte esencial de su conceptualización es integrarlo en los procesos de generación de información estadística, quedando entonces las etapas del proceso como sigue:

- Planeación.
- Diseño conceptual.
- Diseño muestral.
- Definición del procedimiento de captación y procesamiento.
- Captación de la información.
- Procesamiento.
- Almacén de datos (*data warehouse*).
- Presentación de resultados.

Entonces el almacén de datos o *data warehouse* estadístico se incorpora en el proceso convirtiéndose en la fuente de datos única para la presentación de resultados que ahora no se limita a la generación de publicaciones sino que además será la fuente de información para sistemas de consulta interna, desarrollo de productos digitales comercializables y para el sitio del INEGI en Internet.

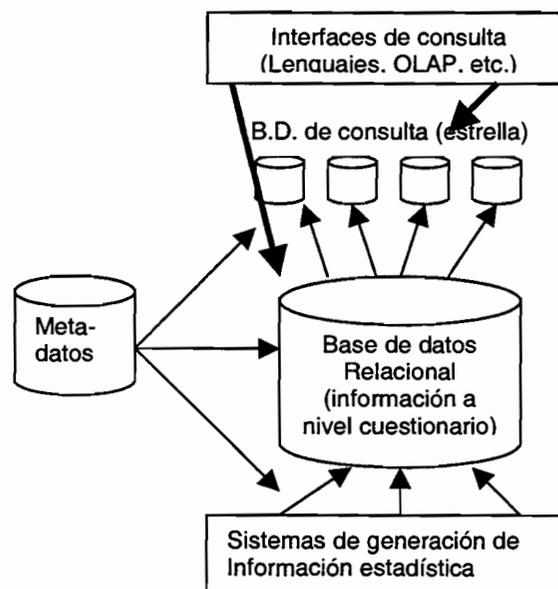
También es de gran utilidad este almacén de datos estadísticos como un insumo para la planeación, diseño conceptual y diseño muestral de subsiguientes eventos estadísticos.

3. Arquitectura de Información

La arquitectura de información institucional considerada en el proyecto se compone de las arquitecturas de datos, de aplicaciones y técnica.

Arquitectura de datos

Define el flujo de información desde los sistemas de captura y procesamiento de la información estadística que a través de los procesos de extracción, transformación y carga se incorporan en la base de datos relacional de información fuente y ésta a su vez es la fuente para información resumizada y/o consolidada. Todo esto apoyado en un repositorio de metadatos.



Arquitectura de aplicaciones

Define las relaciones del proceso del *data warehouse* estadístico con aplicaciones institucionales, entre las que se encuentran:

- Sistemas de procesamiento para generación de información estadística.
- Sistema Integral Administrativo.
- Sitio del INEGI en Internet.
- Sistemas internos de consulta.
- Sitios intranet.
- Productos digitales comercializables.

Arquitectura técnica

Define la infraestructura de hardware que soporta el proyecto, dividida en cinco subsistemas:

- Subsistema de seguridad.
- Subsistema de bases de datos.
- Subsistema de procesamiento.
- Subsistema de aplicaciones.
- Subsistema de almacenamiento.

4. Contenido

El proyecto tiene como objetivo abarcar toda la información estadística que genera el instituto y ser capaz de conectarse a bases de datos generadas por otras instituciones.

Actualmente se tiene incorporados 11 eventos estadísticos en el proyecto:

- XII Censo General de Población y Vivienda, 2000.
- Censos Económicos 1999.
- Estadística de Educación ciclo 1998-1999.
- Encuesta sobre Seguridad y Orden Público 1998 a 2001.
- Estadística de mortalidad 1985 a 2001.
- Estadística de nacimientos 1985 a 2001.
- Encuesta Nacional de Empleo Urbano 2001.
- Encuesta Nacional de Ingreso Gasto de los Hogares 1996, 1998 y 2000.
- Encuesta Industrial Mensual 1998 a 2002.
- Censo de Población y Vivienda 1990.
- Estadística de Comercio Exterior.

Y se tiene planeado continuar esta incorporación de proyectos, entre otros, con los siguientes:

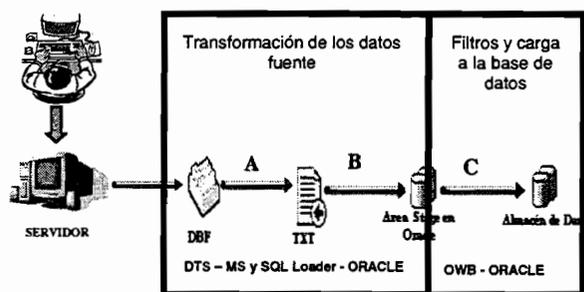
- Censo de Población y Vivienda 1995.

- Censo Ejidal y de Comunidades Agrarias 2001 y 1991.
- Censos Agropecuarios 1991.
- Censos Económicos 1994.
- Estadística de Cultura 1995-2001.
- Estadística de Relaciones Laborales 1990-2001.
- Estadística de Matrimonios y Divorcios 1985-2001.
- Encuesta Nacional de Ingreso Gasto de los Hogares 2002.

5. Desarrollo del Proyecto

Extracción, transformación y carga

Son los módulos que cumplen la función de extraer la información de las fuentes de datos (dispersas y en formatos varios) para cargarla en los modelos de base de datos previamente definidos.



Almacén de datos (*data warehouse*)

Se compone de las bases de datos relacionales que contienen la información fuente (a nivel cuestionario) de los proyectos estadísticos.

Actualmente se cuenta con la información de 11 proyectos estadísticos algunos de ellos con más de cien millones de registros.

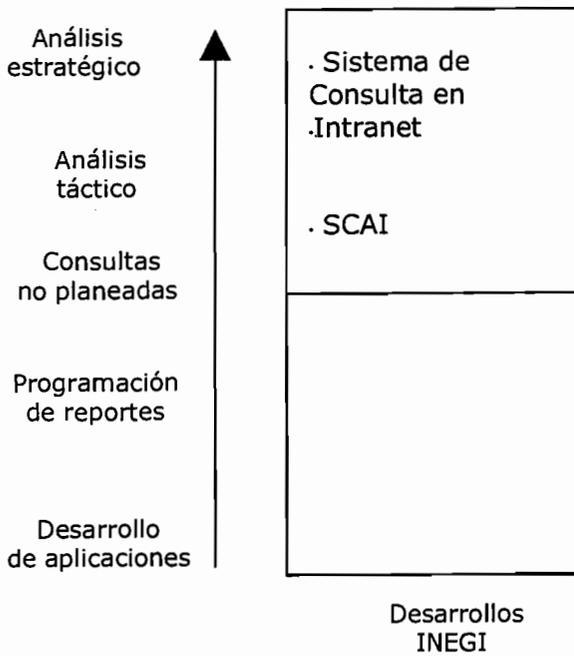
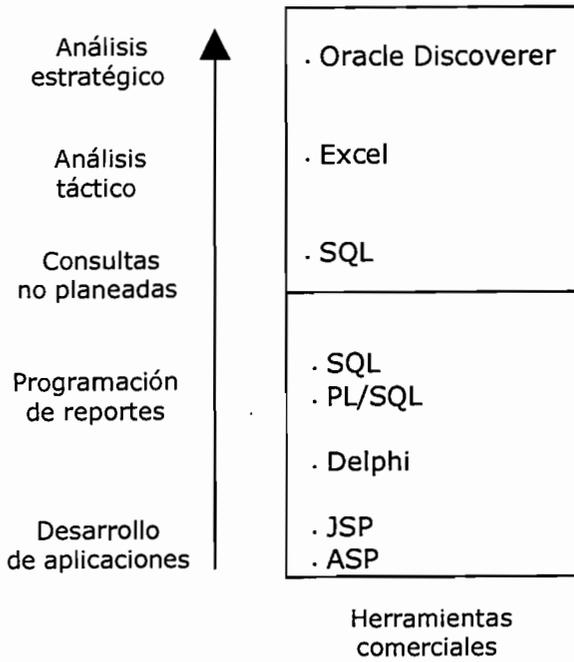
Mercados de datos (*data marts*)

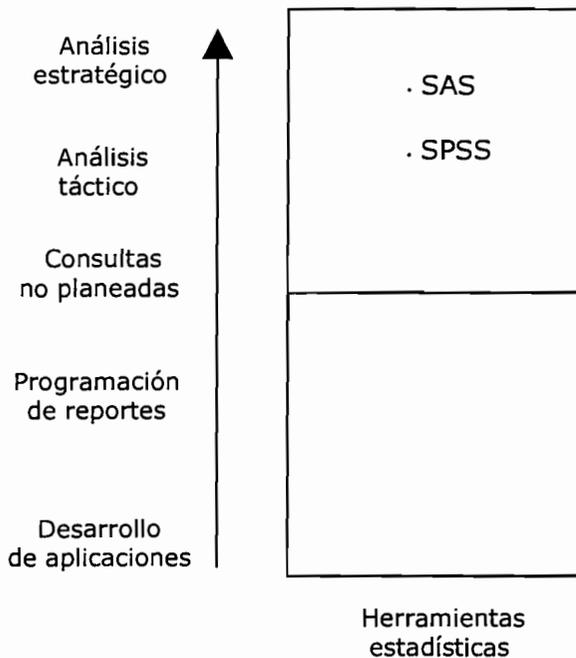
Actualmente se realizan pruebas con diferentes diseños de modelos multidimensionales en esquema estrella para los proyectos de censos de población y económicos.

Se pretende que estos modelos sean el primer contacto de los usuarios finales al consultar la información estadística.

Herramientas de acceso

Debido a que se tiene una amplia diversidad de perfiles de usuario y necesidades de acceso, se ha definido una gama de herramientas disponibles para extraer información de la base de datos.





Metadatos

Por las características del proyecto es necesario poner especial atención en el desarrollo de un subsistema de metadatos que cuente con toda la información necesaria para consultar e interpretar los datos del *data warehouse* estadístico así como relacionar de manera efectiva cada dato con los metadatos correspondientes.

Se han analizado los siguientes estándares internacionales para basar en ellos el subsistema de metadatos:

- ISO-11179.
- Document Data Interchange (DDI).
- Survey Design and Statistical Methodology Metadata (SDSMM), del Buró de Censos de los EEUU.
- Common Warehouse Metamodel.

Este subsistema en desarrollo contará con tres módulos interrelacionados que permiten documentar los datos y los modelos de datos desde diferentes puntos de vista:

- Metamodelo del sistema
- Modelo del negocio
- Registro de elementos de dato

Conclusiones

- El almacén de datos es un proyecto institucional que requiere del compromiso de todas las áreas para poner en marcha los cambios necesarios en la cadena de valor.
- Debe evaluarse en particular la forma en que cada proyecto se incorpora al almacén para mantener la mejor relación costo-beneficio.

- Este proyecto requiere un enfoque general desde su inicio para que la integración sea óptima.
- Lo importante es lo que está detrás de la tecnología: los procesos que soporta para generar valor.
- El data warehouse del INEGI debe proveer la base de información para la prestación de nuevos y mejores servicios en línea que proporcionen a nuestros usuarios mayor capacidad de consulta y análisis para tomar mejores decisiones en beneficio de México.

**ENIGH: VALIDACIÓN Y CONSTRUCCIÓN DE LA
BASE DE DATOS**

BENITO DURÁN

INSTITUTO NACIONAL DE ESTADÍSTICA, GEOGRAFÍA E INFORMÁTICA (INEGI)

MÉXICO

ENIGH: Validación y construcción de la base de datos

Benito Durán

Validación de la encuesta

La validación de la encuesta como una forma de dar calidad a la información comienza en sí, con la captación de la información y continúa con:

- Crítica-Codificación
- Entrada de datos
- Validación de los datos
- Análisis de la información

} Estas tres etapas se realizan en las oficinas de cada una de las entidades de la República Mexicana

Critica-Codificación

Una vez que llega el cuestionario lleno con entrevista completa y correcta, se procede a analizar la información de forma integral, de tal manera que se detecten errores de captación, inconsistencias e incongruencias.

Para llevar a cabo esta etapa se elabora previamente un manual en el que se especifican los pasos y criterios a seguir para que la congruencia y consistencia de la información sea logrado con calidad.

La codificación se realiza después de corregidos los errores e inconsistencias.

Entrada de Datos

Con la finalidad de tener un mayor control y calidad en la entrada de datos se introducen en esta etapa una gran cantidad de criterios de validación, los cuales incluyen: rangos, control de secuencia y congruencia entre variables.

Tomando en cuenta que uno de los objetivos de la entrada de datos es que esta se de de manera veloz y agil solo se introducen aquellos criterios que son cerrados.

Aunque en la etapa de Critica-Codificación se detectan y corrigen una gran cantidad de errores e inconsistencias, en ésta pueden ocurrir otras, es por ello que la misma persona que hace la crítica es la que da entrada a los datos, agilizando con ello esta etapa.

Validación de la información

Se implementa esta etapa como último filtro para garantizar la calidad de la información. Para la ENIGH esta es de gran importancia dado que aquí es donde se detectan la mayoría de los problemas.

La validación de la información se realiza después de haber dado entrada a los datos y se incluye como un módulo adicional del programa de entrada de datos.

Aquí se incluyen tanto congruencias cerradas como abiertas.

Después del ingreso de datos y la validación

Terminada la entrada de datos se procede con:

- Integrar los archivos individuales en uno solo
- Comprobar en el archivo integrado la existencia de la totalidad de los hogares visitados con respuesta completa y correcta
- Transformar el archivo integrado en una base de datos de tablas o archivos temáticos
- Convertir la base de datos al formato del software que se utilizará para el análisis de los datos
- Controlar la consistencia de los datos contenidos en la base de datos. Para ello obtenemos la frecuencia de las variables cualitativas y el valor acumulados de las variables cuantitativas

Construcción de la base de datos

La eficiente manipulación de los datos dependerá de una cuidadosa construcción de la base de datos, tomando en cuenta los puntos siguientes :

- Las tablas o archivos son temáticos
- Cada tabla contiene en sus primeros campos la identificación del hogar o persona
- Los campos en las tablas tienen el mismo orden de las preguntas en el cuestionario
- Los nombres de campo identifican a la pregunta que corresponda en el cuestionario
- Se trata de evitar la redundancia
- Se mantiene la descripción de la base de datos y una clasificación de códigos

Análisis preliminar de la información

Una forma más de validar la información es haciendo un análisis general de la información, para ello se genera un listado de cifras control. Una vez obtenidas las cifras control, la información es comparada con datos históricos y de otras fuentes.

**PROCESAMIENTO DE DATOS DE LA ENCUESTA DE
MEDICIÓN DEL NIVEL DE VIDA**

INSTITUTO NACIONAL DE ESTADÍSTICA Y CENSOS (INEC)

NICARAGUA

ÍNDICE

	<u>Página</u>
Introducción	165
Aspectos Generales	165
Procesamiento de datos	165
Características del sistema de procesamiento de datos	165
Diseño de diccionarios y definición de variables.....	166
Codificación	166
Tipos de inconsistencia	167
Diseños computación de las bases de datos	167
Métodos para la imputación de datos.....	168
Aplicación de los programas de inconsistencias	168
Integración de variables	169
Disposición de los archivos de datos	169
Documentación del sistema.....	172
Generación de tabulados	172
Políticas institucionales de acceso a las bases de datos	172

Introducción

El Instituto Nacional de Estadísticas y Censos realizó la Encuesta sobre Medición de Nivel de Vida en el año 2001, en el marco del Programa MECOVI-Nicaragua, patrocinado por el Banco Mundial, el BID, y el PNUD, y contó además con la contribución económica de los gobiernos de Noruega, Dinamarca y Suecia, así como con fondos de contrapartida del Gobierno de Nicaragua.

La encuesta fue de carácter multipropósito, cubrió 10 grandes temas, y siguió a un conjunto de hogares que ya habían sido entrevistados en la encuesta de similar propósito en 1998. La metodología empleada fue la de las encuestas LSMS del Banco Mundial.

Aspectos Generales

Procesamiento de datos

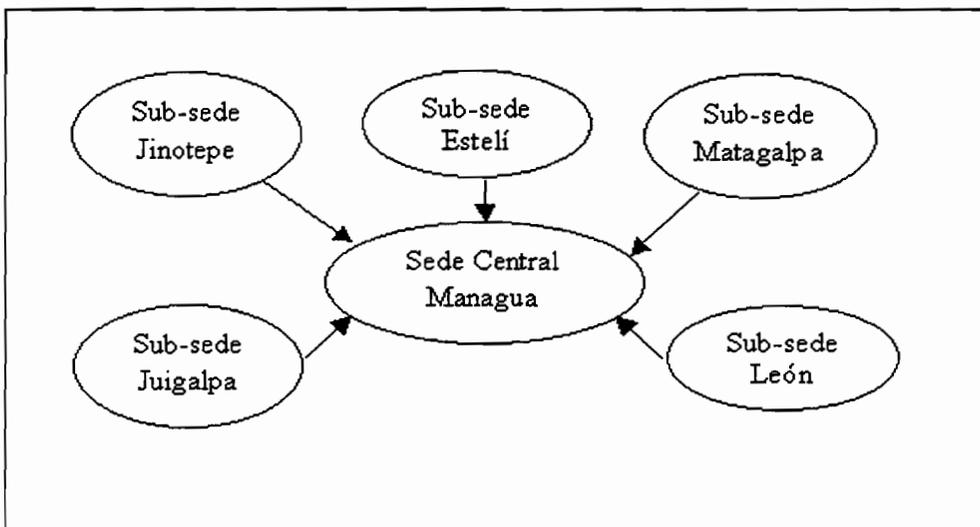
Para el procesamiento de datos de la EMNV, se planteó como meta reducir los errores que se cometieron en 1998, tanto en la digitación como en la implementación de un sistema que reportara los errores procedentes de campo. Esto requería de un software que permitiera desarrollar un sistema inteligente de entrada de datos y de consistencia y que, además, minimizara el tiempo de programación, ya que no se contaba con un tiempo holgado para el desarrollo del sistema. Esto conllevó a pensar en el software ISSA desarrollado por Macro International el cual reunía las características requeridas para desarrollar el sistema que se necesitaba.

Para el desarrollo del sistema se tuvo una buena coordinación entre el equipo técnico de la encuesta y el personal de procesamiento de datos para unificar criterios que conllevaran a garantizar una mejor calidad de los datos.

Características del sistema de procesamiento de datos

La estrategia fundamental fue realizar la consistencia mecanizada en el menor tiempo posible para retroalimentar al personal de campo sobre la tendencia de errores sistemáticos y para asegurar la corrección de los mismos en el campo. Para ello se descentralizó la digitación y consistencia en 6 sedes departamentales, las cuales contaron con el equipamiento y personal suficiente para soportar la carga de trabajo planificada en la operación de campo. Las sedes fueron Jinotepe, Estelí, Matagalpa, Juigalpa, León y Managua. Al final toda la información fue consolidada en la sede central de Managua.

Esquema del sistema de procesamiento de datos



Diseño de diccionarios y definición de variables

Para el desarrollo del sistema de entrada de datos y el control de datos al sistema se emplearon 5 diccionarios:

- i). EMNV2001.dic (diccionario principal) en él están diseñados todos los formatos de entrada de datos y definiciones de variables para el sistema.
- ii). Geografi.dic este diccionario se usaba para hacer una búsqueda en una tabla que contenía la posición geográfica de la vivienda en el segmento censal.
- iii). Workdict.wst este es un tipo de diccionario donde se declaran variables de trabajo de ámbito global que serán usadas en cualquier punto de la aplicación.
- iv). Ant2001.dic diseñado para la captura del cuestionario de antropometría.
- v). Comunit.dic diseñado para la captura del cuestionario comunitario o de precios.

El diseño del diccionario principal fue concebido de tal forma que fuera una réplica exacta del cuestionario del módulo de hogares. Por tal motivo el diccionario principal está dividido en secciones y cada sección contiene un formato de presentación de pantalla el cual contiene todas o parte de las variables de una sección del cuestionario, las definiciones de nombres de variables se hizo de tal forma que su morfología indicara a qué sección y pregunta pertenecía dentro del cuestionario y de esta forma facilitar la programación y posteriormente al procesamiento de datos, un ejemplo de nombre de variable sería la siguiente "S1P1" este nombre de variable se describiría de la siguiente manera:

S1 = sección uno del formulario.

P1 = pregunta 1 de la sección a que se hace referencia.

Codificación

La codificación de las variables del cuestionario que requerían esta acción fue realizada por el supervisor de digitación.

Las secciones y variables a codificar fueron las siguientes:

- i). Carátula. Codificación geográfica (Departamento, municipio, comarca y localidades).
- ii). Sección 4. Temas de capacitación.
- iii). Sección 5. Rama de actividad, ocupación y codificación geográfica (Departamento, municipio y país).
- iv). Sección 6. Codificación geográfica (Departamento, municipio y país) y Ocupación de la persona.
- v). Sección 8. Rama de actividad del negocio del hogar.
- vi). Sección 9. Unidades de medidas de los productos de consumo básico.
- vii). Sección 10. Códigos de árboles, cultivos y unidades de medidas aplicadas a los cultivos y productos de origen animal.

Para la codificación se emplearon los siguientes instrumentos:

- i). Clasificador de Ocupaciones.
- ii). Clasificador de Rama de Actividad.

- iii). Clasificador de Departamento y Municipio (DPT).
- iv). Clasificador de Países.
- v). Clasificador de Unidades de Medidas según Producto o Cultivo.
- vi). Clasificador de Cultivos, árboles frutales y árboles no frutales.
- vii). Clasificador de estado del cultivo.
- viii). Clasificador de temas de capacitación.
- ix). Clasificador de comarcas y localidades.

Tipos de inconsistencias

Para la programación se tomaron en cuenta los siguientes tipos de inconsistencias:

- i). Rangos en variables cerradas: en aquellas variables que tenían un rango definido en el cuestionario, se controlaba que las respuestas tuvieran un rango de valor permisible establecido en el diccionario de variables.
- ii). Rangos en variables abiertas:
 - a) Variables monetarias. Ellas eran revisadas por el equipo de análisis el cual generaba un reporte de valores altos o bajos según su criterio.
 - b) Variables codificadas. Estas eran verificadas por el sistema de captura de datos dentro de una tabla de códigos definida dentro del sistema.
- iii). Flujos: Eran controlados por el sistema de captura de datos para agilizar la digitación y evitar que el digitador introdujera datos en variables que no correspondieran
- iv). Lógicas: Consistía en comparar preguntas que tuvieran relación entre ellas, sea en la misma sección o entre secciones diferentes.
- v). Datos faltantes: Estos eran reportados por el supervisor de digitación o el digitador anotando observaciones en el cuestionario. Esto se detectaba al momento de la digitación ya que no se permitía valores en blanco en los campos que deberían tener un valor.

Diseños computacionales de las bases de datos.

El software que se utiliza para la entrada de datos, y validación es el ISSA-CSPPro creado por Macro International debido a su agilidad para el control de los flujos de datos. El diseño y programación del sistema estuvo a cargo del responsable de procesamiento de la encuesta. En principio el objetivo principal era diseñar un sistema que disminuyera los errores de digitación, para esto se programaron una serie de condiciones que emitieran mensajes de advertencia al digitador para efectuar las revisiones que fueran necesarias.

Las pantallas de captura de datos se diseñaron de tal forma que el digitador tuviera acceso visual a todas las variables del formato que se estaban digitando en ese momento, esto permitía al digitador ubicarse rápidamente en la sección y pregunta del cuestionario.

En esta parte del diseño se definió que el sistema controlara todos los flujos que presentaba el cuestionario, con el objetivo de agilizar la digitación de datos y evitar que el digitador introdujera datos en preguntas que no aplicaban, también se definieron variables de control adicionales a las establecidas en el cuestionario, para ayudar al sistema a controlar la entrada de datos.

Considerando que habían diferentes tipos de cuestionarios (hogares, antropometría y comunitario), se elaboró un menú de selección para dirigir al digitador hacia el tipo de formulario que debía ingresar al sistema, y para que se activaran los diccionarios de variables correspondientes.

Además de los programas de captura de datos para los diferentes cuestionarios y el programa de consistencia, se tuvo que elaborar una serie de programas con los que se exportaban los datos a un formato entendible para la mayoría de los usuarios y por ende a su procesamiento y análisis estadístico, para esto se diseñaron 46 programas de exportación de datos.

Las pruebas al sistema de captura de datos se hizo al inicio con cuestionarios que se levantaron en las pruebas pilotos, con el objetivo de comprobar los resultados del programa. En esta fase se involucró a los digitadores para que fueran conociendo el sistema con el cual iban a trabajar. Ello sirvió para evaluar el sistema ya que se reportaron errores y rupturas del programa, para la modificación de los programas y ponerlos a punto para el momento de iniciar la digitación de los datos reales de la encuesta.

Métodos para la imputación de datos.

Las reglas de inconsistencia fueron preparadas por el equipo técnico de metodología de la encuesta quienes definieron las relaciones lógicas entre preguntas que estaban concatenadas entre sí, en la misma sección o entre secciones diferentes del cuestionario; un ejemplo de este tipo de inconsistencia es el siguiente: si la persona respondía en la sección 4 que no se había matriculado en el año lectivo actual, era inconsistente que respondiera en la sección 5 que la razón de no trabajar era por estudio.

La consistencia de flujos de preguntas que no aplicaban según las respuestas a preguntas anteriores las controlaba el sistema de entrada de datos. Estos eran casos tales como saltos de una pregunta a otra y flujos de información de personas que según edad o sexo no aplicaban a una sección, o una serie de preguntas dentro de una sección. Esta parte del control de inconsistencias fue preparada por el diseñador del sistema de entrada de datos.

Se estableció un método de codificación para los datos: el carácter "." Corresponde a las variables que no aplicaban según el flujo del cuestionario. Cuando se presentaban casos omitidos por el encuestador o por el informante se estableció el código "9", "99", "999". etc., esto dependiendo de la longitud del campo de la variable.

Cabe señalar que parte de las inconsistencias fueron integradas al programa de captura de datos, esto se hizo con el objetivo de corregir o evitar errores en tiempos de digitación. Ej.: Fecha de nacimiento de la persona, el sistema rechazaba si la edad declarada no correspondía a la fecha de nacimiento.

En estos programas se generaron funciones de consultas a tablas de códigos especiales para cada tipo de variable. Ellas fueron:

- i). Consultas a códigos de rama de actividad y ocupación.
- ii). Consulta a códigos geográficos (departamento, municipios y países).
- iii). Consulta de unidades de medidas según producto o cultivo.
- iv). Consulta a códigos de cultivos y árboles.

Aplicación de los programas de inconsistencias

El programa de detección de inconsistencias fue elaborado en el software ISSA que contiene una opción llamada "batch", la cual permite evaluar los datos grabados en una base de datos, generando un archivo de salida de las inconsistencias halladas.

Este programa se fue actualizando durante la primera etapa del levantamiento de

la encuesta con nuevas reglas que se iban incluyendo conforme se detectaran errores, y se ponía en funcionamiento en todas las sedes regionales de digitación.

El programa de inconsistencias era aplicado por el supervisor de digitación a cada conglomerado en las sedes regionales de digitación; los errores detectados eran listados y luego revisados y discutidos con el supervisor de brigada y/o el crítico. Esto conllevaba a revisar el cuestionario para determinar la solución de la inconsistencia; si la inconsistencia era solucionable en oficina, se hacía la corrección, de lo contrario el cuestionario era enviado a campo para la recuperación del dato.

Los datos faltantes por omisiones del enumerador o por ausencia del informante eran reportados a los supervisores de brigadas para su recuperación en campo, estos reportes eran elaborados de una forma manual y consistían en observaciones anotadas por el digitador en el cuestionario.

Una vez corregido el cuestionario, el digitador efectuaba la corrección en el archivo de datos y se aplicaba nuevamente el programa de inconsistencias para verificar si se había realizado el cambio y para detectar otras inconsistencias que pudieran haberse presentado con el cambio. Este proceso era repetido entre una a tres veces en la mayoría de los casos.

La alimentación de información a la sede central se hacía al finalizar cada ronda de trabajo por cada supervisor de digitación. Los archivos eran entregados al responsable de procesamiento, quien a su vez volvía a aplicar el programa de inconsistencia a los datos de cada sede. En esa etapa se generaron reportes de inconsistencias por sede de digitación para su revisión, aclaración y posterior corrección.

Una vez que la información era revisada y corregida se remitía al grupo de analista del área de los índices de pobreza para su evaluación. Este equipo desarrolló pruebas para detectar valores extremos en las variables que iban a constituir el agregado de consumo y las variables de ingresos (salvo la sección agropecuaria); asimismo, elaboró tabulaciones cruzadas entre variables para detectar inconsistencias o falta de información, o en las frecuencias. Los errores detectados en esta etapa obligaron a una nueva revisión de los cuestionarios, y eventualmente al retorno a campo. Luego de cada corrección se aplicaba nuevamente el programa global de consistencia.

La no respuesta parcial así como los hogares que no completan la segunda fase se eliminan de la base de datos.

Integración de variables.

Las variables de consumo son las únicas que se integran a la base de datos tras una proceso posterior.

Disposición de los archivos de datos

Existen tres tipos de archivos: uno de datos originales que refleja el contenido de los cuestionarios tales como vinieron de campo en la primera oportunidad, en formato ISSA, de igual manera los datos corregidos tenían el mismo tipo de formato y los datos finales en formato SPSS.

Las bases de datos originales están en formato ASCII o archivos planos que solo pueden ser leídos o entendidos disponiendo del diccionario creado en ISSA. Se crearon un total de 427 archivos, cada uno corresponde a una base de datos de un conglomerado que correspondía a la información de 10 cuestionarios de hogares en el área rural y de 12 en el área urbana. Este criterio fue establecido con el fin de proteger la información por el motivo que si se dañaba algún archivo solo se perdía información del conglomerado correspondiente.

Los datos finales para usuarios están disponibles en formato SPSS y están desagregados en 50 tablas, cada tabla contiene el rótulo de cada variable y sus respectivas categorías de respuestas, para facilitar la comprensión de los datos que se encuentran grabados en dicha tabla. Las tablas que componen la base de datos de nivel de vida del 2001 son las siguientes:

1. EMNV01 VIVIENDA.SAV : Contiene las variables de identificación de cada hogar y la sección 1 del formulario de hogares, que va de la preg. 1 a la preg. 58.
2. EMNV02 VIVIPROGRAMAS.SAV: Contiene información de los programas de beneficio social que han beneficiados a los miembros del hogar esta parte es continuación de la sección 1 y comprende de la preg. 59 a la preg. 62.
3. EMNV03 POBLACION.SAV: Contiene información de cada una de los miembros, no miembros y ausentes del hogar, esta tabla contiene la sección 2 hasta la sección 7 excluyendo solamente la parte de migración externa y mortalidad.
4. EMNV04 MIGRACION.SAV: Información de los familiares en el extranjero que declaro el hogar, sección 6 parte B.
5. EMNV05 MORTALIDAD.SAV: Información de los familiares fallecidos que el hogar declara, sección 7 parte C.
6. EMNV06 NEGOCIO PARTE A.SAV: Contiene información de los negocios que pertenecen al hogar, Sección 8 parte A.
7. EMNV07 NEGOCIO PARTE B.SAV: Gastos del negocio que se realizaron en los últimos 12 meses, Sección 8 parte B.
8. EMNV08 NEGOCIO PARTE C.SAV: Capital e inventario del negocio, Sección 8 parte C.
9. EMNV09 GASTOS PARTE_A.SAV: Contiene información de todas las variables simples que pertenecen a la sección de gasto del hogar o sección 9.
10. EMNV10 GASTOS PRODUCTOS.SAV: Productos básicos de consumo del hogar adquiridos en los últimos 15 días, preg. 3 a preg. 10 sección 9 parte A.
11. EMNV11 GASTOS PARTE B1.SAV: Gastos del hogar en la semana pasada, sección 9 parte B1.
12. EMNV12 GASTOS PARTE B2.SAV: Gastos del hogar en el mes pasado, sección 9 parte B2.
13. EMNV13 GASTOS PARTE B3.SAV: Gastos del hogar en los últimos 6 meses, sección 9 parte B3.
14. EMNV14 GASTOS PARTE B4.SAV: Gastos del hogar en los últimos 12 meses, sección 9 parte B4.
15. EMNV15 GASTOS PARTE C.SAV: Inversiones echas por el hogar durante los últimos 12 meses, sección 9 parte C.
16. EMNV16 GASTOS PARTE D1.SAV: Ingresos del hogar en el mes pasado, sección 9 parte D1.
17. EMNV17 GASTOS PARTE D2.SAV: Ingresos del hogar en los últimos 12 meses, sección 9 parte D2.
18. EMNV18 GASTOS PARTE D3.SAV: Remesas recibidas por el hogar en los últimos 12 meses, sección 9 parte D3.
19. EMNV19 GASTOS EQUIPO.SAV: Equipamiento del hogar sección 9 parte E.
20. EMNV20 GASTOS EQUI_EXT.SAV: Equipamiento adicional del hogar sección 9 parte E.
21. EMNV21 AGRO VARIABLES SIMPLES.SAV: Contiene todas las variables simples que pertenecen a cualquier parte de la sección 10.
22. EMNV22 AGRO FINCAS PROPIAS.SAV: Fincas o parcelas propiedad de algún miembro del hogar, sección 10 parte A. Preg 6 a preg. 21.

23. EMNV23 AGRO FINCAS ALQUILADAS.SAV: Fincas o parcelas alquiladas por algún miembro del hogar, sección 10 parte A. Preg 22 a preg. 32.
24. EMNV24 AGRO COMPRA Y VENTA DE FINCAS.SAV: Fincas o parcelas vendidas o compradas en los últimos 5 años sección 10 parte A, preg. 34 a preg. 39.
25. EMNV25 AGRO PARTE B.SAV: Producción forestal sección 10 parte B, preg 41. a preg. 43.
26. EMNV26 AGRO PARTE C1.SAV: Actividad agrícola sección 10 parte C1. producción, preg. 44 a preg. 52.
27. EMNV27 AGRO PARTE C2.SAV: Actividad agrícola sección 10 parte C2. insumos, preg.53 a preg. 56.
28. EMNV28 AGRO PARTE C3.SAV: Actividad agrícola sección 10 parte C3. venta de futuro o adelantada, preg. 57 a preg. 61.
29. EMNV29 AGRO PARTE D.SAV: Actividad Pecuaria sección parte D1, D2 y D3, preg. 70 a preg. 82.
30. EMNV30 AGRO PARTE E PRODUCTOS.SAV: Producto y subproductos de origen agrícola y animal sección 10 parte E, preg. 90 a preg.94.
31. EMNV31 AGRO PARTE E SUBPRODUCTOS.SAV: Producto y subproductos de origen agrícola y animal sección 10 parte E, preg. 95 a preg.97.
32. EMNV32 AGRO FUERZA DE TRABAJO SIN ALIMENTACION.SAV:
33. Fuerza de trabajo sin alimentación, sección 10 parte F1. preg. 99.
34. EMNV33 AGRO FUERZA DE TRABAJO CON ALIMENTACION.SAV: Fuerza de trabajo con alimentación, sección 10 parte F1. preg. 100.
35. EMNV34 AGRO FUERZA DE TRABAJO PERMANENTE.SAV: Fuerza de trabajo permanente, sección 10 parte F1. preg. 102.
36. EMNV35 AGRO FUERZA DE TRABAJO DE MIEMBROS.SAV: Fuerza de trabajo de los miembros del hogar, sección 10 parte F1. preg. 103.
37. EMNV36 AGRO PARTE F2.SAV: Gastos en actividades agropecuarias y forestales, sección 10 parte F2, preg. 105 a preg. 106.
38. EMNV37 AGRO EQUIPOS AGROPECUARIOS.SAV: Equipo agropecuario, sección 10 parte G1. preg. 107 a preg. 110.
39. EMNV38 AGRO INSTALACIONES AGROPECUARIAS.SAV: Instalaciones agropecuarias, sección 10 parte G2. preg. 111 a preg. 114.
40. EMNV39 AGRO PRODUCCIÓN DE PATIO AGRÍCOLA.SAV: Producción de patio agrícola, sección 10 parte I1. preg 143 a preg. 148.
41. EMNV40 AGRO PARTE I2.SAV: Insumos, sección 10 parte I.2, preg. 149 a preg. 150.
42. EMNV41 AGRO PARTE J1 Y J2.SAV: Producción de patio pecuario, sección 10 parte J.1, preg. 152 a preg. 162.
43. EMNV42 AGRO PARTE K.SAV: Productos y subproductos de la producción de patio, sección 10 Parte K. Preg. 163 a preg. 166.
44. EMNV43. AGRO PARTE K SUB PRODUCTOS.SAV: Productos y subproductos de la producción de patio, sección 10 Parte K. Preg. 167 a preg. 168.

45. EMNV44 AGRO FUERZA DE TRABAJO DE PATIO.SAV: Fuerza de trabajo utilizada en la producción de patio, sección 10 parte L. Preg. 169.

Estas 45 tablas tienen una clave común que corresponde al número de formulario de cada hogar, el cual es único, lo cual les permite relacionarse entre si. Las variables que contienen el número del formulario son las I00A y I00B.

Documentación del sistema

La información sobre el sistema utilizado en la digitación y procesamiento de datos está contenido en:

1. Manual del digitador: Comprende el funcionamiento del sistema mediante un menú de selección que domina todas las entradas de datos de los diferentes cuestionarios.
2. Manual instructivo básico de ISSA: Empleado para la capacitación de los supervisores de digitación.
3. Programas fuentes y diccionarios usados en el sistema.

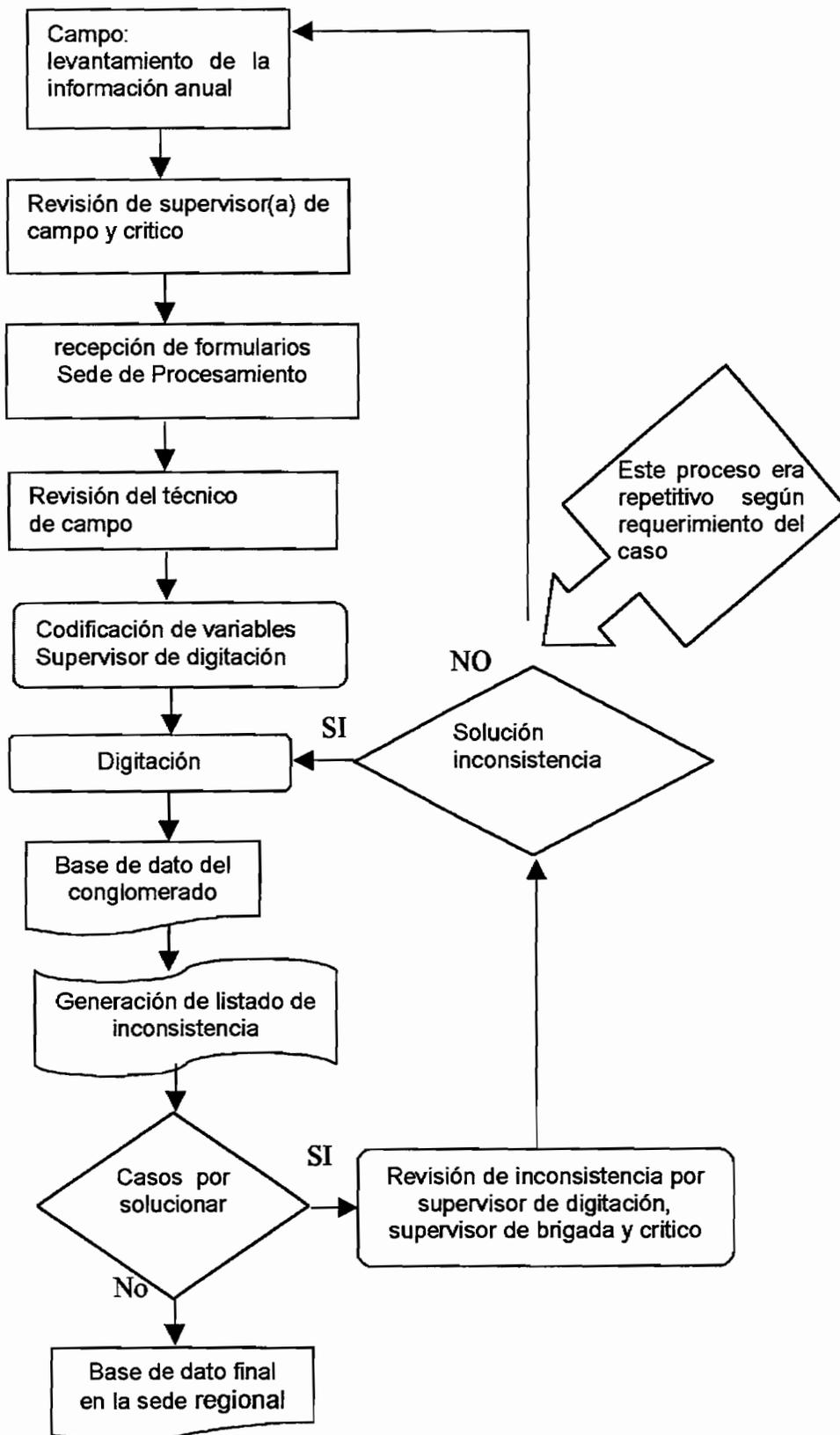
Generación de tabulados.

Los tabulados son generalmente trabajados con SPSS, no se han generado sistemas propios para la generación de estos.

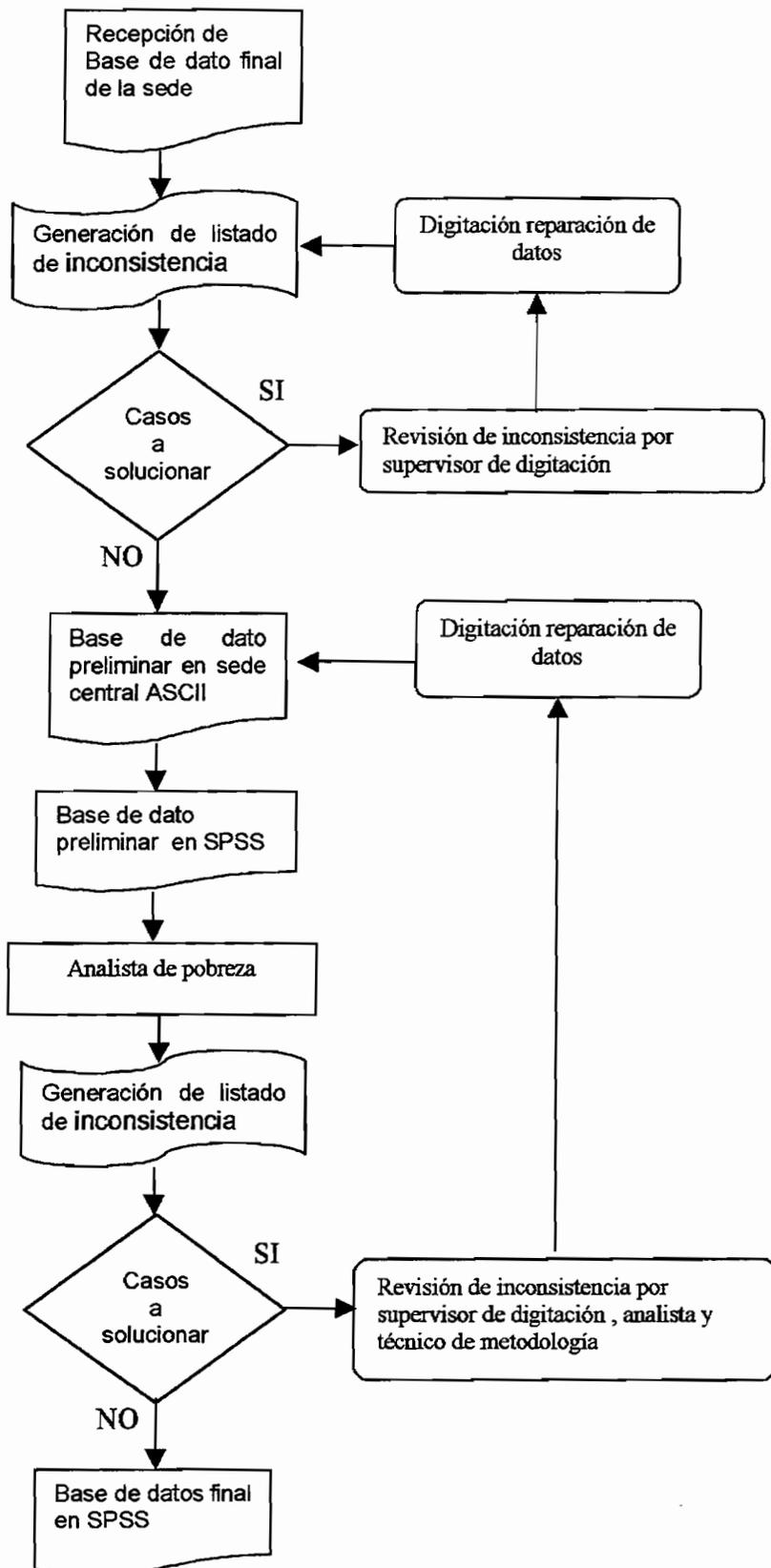
Políticas institucionales de acceso a las bases de datos.

No existe un calendario preestablecido para la entrega de resultados. El acceso a las bases de datos se entregan a los principales usuarios sin consto, las tablas de resultados están en la página web del INEC y los procesos en línea están aún en desarrollo.

Esquema de recepción de datos en sede regional



Esquema de recepción de datos en sede central:



**INFORME METODOLÓGICO INFORMÁTICO
ENCUESTA DE HOGARES AÑO 2002**

MARCELINO PEÑALBA GARCÍA

DIRECCIÓN DE ESTADÍSTICA Y CENSO (DEC)

PANAMÁ

ÍNDICE

	<u>Página</u>
Introducción	177
Aspectos Generales	177
Procesamiento de Datos	181

Introducción

Para facilitar el desarrollo de los diferentes Sistemas, se dispone de la siguiente documentación: listado de empadronadores y supervisores, por provincia y código; listado para la recodificación de la condición de actividad; listado de validaciones e inconsistencias; manual del encuestador; cuestionario final; archivo con los segmentos que conforman la cobertura de campo de la encuesta y archivo con los segmentos indígenas.

El cuestionario tiene un total de 49 preguntas repartidas en 6 capítulos. En su mayoría las preguntas tienen códigos preimpresos. La codificación propiamente dicha se da en preguntas como la Ocupación, Rama de Actividad de la Empresa, Ingreso, Lugar Geográfico del Trabajo y algunas otras. Los capítulos VII y VIII son utilizados por instituciones del Gobierno que desean utilizar el vehículo de la encuesta de Hogares para hacer investigaciones cortas. Instituciones tales como el Ministerio de Economía y Finanzas, Ministerio de la Juventud, la Niñez, la Mujer y La Familia, entre otros.

Aspectos Generales

Para el Sistema de la encuesta de Hogares se desarrollaron 5 módulos (Revisión del año 2001), a saber: Módulo de Captura de Datos, Módulo de Depuración de Cobertura, Módulo de Recodificación, Módulo de Tabulados, Módulo Factor de Expansión. El Módulo de Captura de Datos en su inicio fue desarrollado utilizando el paquete de Visual Foxpro (versión 2.6), el cual está dividido en 4 niveles: Adición, Consulta, Modificación y Eliminación. A su vez el nivel de Adición está dividido en los programas de: Standard, Persona, Anexo1 y Anexo2. Los niveles de Consulta, Modificación y Eliminación no utilizan el programa Standard. Las validaciones e inconsistencias se corrigen en línea. Esto quiere decir que el Sistema no le permite al capturador continuar, a menos que haya hecho las correcciones debidas. Para mayor seguridad en la integridad de los datos, una vez finalizado el período de la captura, se hacen verificaciones por lote para ciertas variables importantes.

En años anteriores se formaba una base de datos, cuyos registros eran incluidos cada vez que se disparaba una inconsistencia u error en la validación. No era fácil identificar si el error era producido por captura o por codificación de las preguntas. Lo importante era detectar qué preguntas estaban dando mayor dificultad para así darles seguimiento durante la revisión, codificación y captura.

Módulo de Captura de Datos: Se desarrolló bajo la arquitectura de Cliente-Servidor (para la sede en Contraloría). Esto quiere decir que el sistema en modo ejecutable está en el Servidor al igual que sus diferentes componentes (38 archivos), incluyendo las 3 bases de datos que reciben registros de información de acuerdo al contenido de los cuestionarios. Todas las tardes se realiza un backup (respaldo) del Servidor hacia la PC del Administrador de los Sistemas Informáticos de la encuesta de Hogares. En el interior de la República existen Oficinas Regionales de la Dirección de Estadística y Censo. En cuatro de ellas se concentra la captura de datos. Se les atiende por medio de gira de trabajo para instalar el sistema en los servidores y dar algunas instrucciones. Luego de dos meses se regresa para depurar la cobertura y traernos en disquete las bases de datos.

Los capturadores que se utilizan son personas que por lo general tienen experiencia de campo en la antes mencionada encuesta. Para la Dirección de Estadística y Censo todo dato que se le solicite a un ciudadano es materia confidencial. Por ende para el capturador el cuestionario y todo lo relacionado con el mismo es material sensible. Existe un control para determinar qué capturador está capturando qué carpeta (de 4 a 6 unidades primarias) y desde qué fecha. El cuestionario no puede ser sacado del área de trabajo.

La verificación inicial de la captura de un cuestionario recae sobre una base de datos llamada Cobertura, la cual verifica que esa unidad primaria sea válida (Provincia + Estrato + Distrito + Corregimiento + Unidad). Si lo anterior es correcto, se continúa ingresando los campos de Cuestionario + Hogar +

Número de Persona. Éstos unidos a los de la Cobertura, verifican en las otras 2 base de datos para que no exista duplicidad en las llaves. Se capturaron un total de 13,404 cuestionarios, agosto de 2002. Quedando los datos distribuidos en las 3 bases de datos (antes mencionadas), de la siguiente manera:

Hogar.dbf – 12.0 MB con 54,500 registros
 Anexo1.dbf – 2.90 MB con 13,404 registros
 Anexo2.dbf – 3.32 MB con 54,500 registros

Diseño del registro de la encuesta de Hogares-agosto '2002 (solo la base de Hogar), levantada en base de datos (Visual Foxpro v 2.6):

CAMPO	TIPO	TAMAÑO	DESCRIPCION
PROV	C	2	Provincia
DIST	C	2	Distrito
CORRE	C	2	Corregimiento
ESTRA	C	2	Estrato
UNIDAD	C	3	Unidad primaria
CUEST	C	2	Cuestionario
HOGAR	C	1	Hogar
NPER	C	2	Número de la persona
SEMANA	C	1	Semana en que se encuestó
CAPTURA	C	2	Capturador
EMPA	C	3	Empadronador/Encuestador
SUPER	C	2	Supervisor/Coordinador
FECHA	DATE	8	Fecha de captura
NUCLEO	C	2	Núcleo Familiar
P1	N	1	Parentesco
P2	N	1	Sexo
P3	N	2	Edad
P4	N	1	¿Cuál es su estado conyugal actual?
P5	N	1	¿Tiene ud. Seguro Social?
P6	N	3	¿Dónde residía en agosto de 2001?
P7	N	1	¿Asiste a la escuela actualmente?
P7A	N	1	¿Cuál es el motivo principal por el que no asiste a la escuela?
P8	N	2	¿Qué grado o año escolar más alto aprobó?
P9	C	4	¿Qué diploma o título tiene?
P10_17	N	2	¿Trabajó la semana pasada?
P18	N	1	¿Buscó trabajo el mes pasado?
P19	N	1	¿Buscó trabajo durante los últimos tres meses?
P20	N	3	¿Qué gestión hizo...?
P21	N	1	¿Cuánto tiempo hace...?
P22	N	1	¿Piensa buscar trabajo...?
P23	N	2	¿Por qué motivo no estuvo buscando...ni piensa buscar?
P24	N	1	¿Para qué tipo de trabajo está disponible..?
P25	N	3	¿Cuánto tiempo hace que realizó su último trabajo?

CAMPO	TIPO	TAMAÑO	DESCRIPCION
P26	C	4	¿Qué ocupación oficio o trabajo realizó....?
P27	N	1	¿Dónde trabajó?
P28	C	4	¿A qué se dedica ese negocio, establecimiento...?
P29	N	1	¿Cuántas personas trabajan en el negocio...?
P30	N	1	¿Dónde ud. Trabaja o trabajó por última vez...?
P31	N	1	¿Es o era empleado..?
Pregunta # 32:			...Sitio habitual de trabajo:
P32A	C	2	Provincia
P32B	C	2	Distrito
P32C	C	2	Corregimiento
P33	C	3	¿Qué tiempo tiene de trabajar en ese negocio...?
Pregunta # 34:			
P34	N	1	¿Trabajó ud. Por sueldo fijo...? Sí = 1/No = 2
P341	N	4	Por día?
P342	N	4	Por tarea?
P343	N	4(2 dec.)	Por hora?
P344	N	2	Días
P345	N	2	Tareas
P346	N	3	Horas
Pregunta # 35:			
P351	N	4	Salario en efectivo
P352	N	4	En especie
P353	N	4	Ingreso en efectivo...
P36	N	2	¿Cuántas horas trabajó la semana pasada en la...?
P37	N	1	¿Tuvo algún otro trabajo la semana pasada...?
P38	C	4	¿...Ocupación otro trabajo?
P39	C	4	¿A qué se dedica ese negocio...?
P40	C	3	¿Qué tiempo tiene de trabajar en ese negocio...?
P41	N	2	¿Cuántas horas trabajó la semana pasada...otro trabajo
P42	N	4	¿Ingreso mensual...otro trabajo?
P43	N	1	¿Deseaba trabajar más horas
P44	N	1	¿Por qué no trabajó más...
P45	N	1	¿Buscó trabajo adicional...
P46	N	1	¿...Trabajo adicional disponible...?
Pregunta # 47:			
P47A	N	4	Jubilación o pensión?...
P47B	N	4	Ayuda familiar?...
P47C	N	4	Alquileres, rentas...?...
P47D	N	4	Premios de lotería...?...
P47E	N	4	Becas o subsidios?...
P47F	N	4	Décimo tercer mes?...

CAMPO	TIPO	TAMAÑO	DESCRIPCION
P47G	N	4	Ingresos agropecuarios?...
P47H	N	4	Otros ingresos?...
P47I	N	1	Sin ninguno de estos ingresos?...
Los campos que a	Continuación	Aparecen es	Cuando Nper = "01" y P1 = 1
PERSONAS_H	N	2	Total de personas, sexo masculino.
PERSONAS_M	N	2	Total de personas, sexo femenino.
PERSONAS	N	2	Total de personas.
E10YMAS_H	N	2	Total de personas de 10 y más años, sexo masculino.
E10YMAS_M	N	2	Total de personas de 10 y más años, sexo femenino.
E10YMAS	N	2	Total de personas de 10 y más años de edad.
E15YMAS_H	N	2	Total de personas de 15 y más años, sexo masculino.
E15YMAS_M	N	2	Total de personas de 15 y más años, sexo femenino.
E15YMAS	N	2	Total de personas de 15 y más años de edad.
MATRI	N	1	Matriculados en 1er. Grado
G4TO	N	1	De ellos ¿cuántos han aprobado 4to. Grado?
Los que no han aprobado, ¿cuál ha sido el motivo?			
FRACASO	N	1	Fracasó
RECURSO	N	1	Falta de recursos
TRABAJAR	N	1	Tiene que trabajar
INTERES	N	1	Falta de interés
ENFERMED	N	1	Enfermedad
OTROS	N	1	Otros motivos

CAMPOS	DE	PROCESAMIENTO
REGIRECO	C	1 Regiones (1,2,3,4)
AREARECO	C	1 Urbano = "1" / rural = "2"
OCU_RECO	C	2 Condición de actividad. Códigos del 01 al 29
RANGO15	C	1 15 y más = 1; menos 15 = 2
RANGO10	C	1 10 y más = 1; menos 10 = 2
FAC15_E	N	3 Factor de expansión de 15 y más/menos 15 años de edad.
FAC10_E	N	3 Factor de expansión de 10 y más/menos 10 años de edad.
RAN10_14	C	1 10 a 14 años = 1 / resto = 2
FAC10_14	N	3 Factor de expansión de 10 a 14 años de edad
RANGO18	C	1 18 y más = 1; menos 18 = 2
U_P	C	7 Unidad primaria: prov + estra + unidad
INDI_REC	C	1 No indígena = 1/Indígena = 2
Total		227 (2 dec.)

Vale la pena mencionar, que la base de datos de la encuesta de Hogares como producto del procesamiento de datos es única y no se generan versiones por correcciones para evitar posibles errores en la explotación de los datos. Los usuarios intermedios y finales, según su necesidad pueden crear otros campos recodificados. Su acceso es restringido al público, proporcionándola sólo a algunas instituciones gubernamentales u organismos internacionales, previamente identificados. Hasta el momento la base de datos de Hogares no se dona, ni se vende a la empresa privada.

Procesamiento de Datos:

Cabe señalar que las directrices de procesamiento de datos que se siguen son las consensuadas con la Dirección Nacional de Informática, las cuales son revisadas todos los años. La ejecución del Censo de Población y Vivienda, el cual se realiza cada 10 años, sirve de marco conceptual para todo lo que es procesamiento de datos dentro de la Dirección de Estadística y Censo.

A continuación se detallan los otros Módulos desarrollados:

Módulo de Depuración de Cobertura: Se desarrollaron programas para revisar la cobertura a nivel de unidad primaria y a nivel de cuestionarios. En ambos casos se confrontaban las cifras obtenidas por el recorrido que hacía el empadronador versus lo capturado, detectándose omisiones y/o sobregistros. El producto del proceso antes mencionado, era vertido en dos listados a ser corregidos o revisados por el personal de la Sección de Población y Vivienda, ubicando los cuestionarios aludidos y haciendo las correcciones necesarias con el Módulo de Captura. Al final de la captura se generó un listado a nivel de República. Cuando toda la cobertura estuvo depurada (todas la Oficinas Regionales más la sede central) se generó un listado con niveles de rompimiento por Estrato, Provincia y República.

Módulo de Recodificación: Se agregaron 12 nuevas columnas (o campos) a las bases de datos, que mediante programación obtienen nuevos valores a partir de variables capturadas. Por ejemplo: upm, indígena, ocupación, rama de actividad, rango de edades, etc. La generación de estos campos recodificados serán utilizados por los dos Módulos que vienen a continuación.

Módulo de Tabulados: Se generan 4 tabulados controles (sólo para la base de Hogar) sin expandir para ver la consistencia de los datos. Una vez expandidas las 3 bases de datos (Hogar, Anexo1 y Anexo2), se generan nuevamente los antes mencionados tabulados. Se utiliza el software Excel (versión Office Profesional, año 1997). A continuación se detallan los tabulados:

Sin y con factor de expansión:

Tabulación 1a. VIVIENDAS OCUPADAS Y POBLACIÓN POR TIPO DE ACTIVIDAD, SEGÚN REGIÓN, ESTRATO Y UNIDAD PRIMARIA EN LA REPÚBLICA: ENCUESTA DE HOGARES, AGOSTO 2002(P). [Para las edades de 15 y más años de edad. Ocupados y Desocupados que trabajaban antes.]

Tabulación 1b. VIVIENDAS OCUPADAS Y POBLACIÓN POR TIPO DE ACTIVIDAD, SEGÚN REGIÓN, ESTRATO Y UNIDAD PRIMARIA EN LA REPÚBLICA: ENCUESTA DE HOGARES, AGOSTO 2002(P). [Para las edades de 15 y más años de edad. Nunca han trabajado e inactivos puros.]

Tabulación 1c. VIVIENDAS OCUPADAS Y POBLACIÓN POR TIPO DE ACTIVIDAD, SEGÚN REGIÓN, ESTRATO Y UNIDAD PRIMARIA EN LA REPÚBLICA: ENCUESTA DE HOGARES, AGOSTO 2002(P). [Para las edades de 15 y más años de edad. Potencialmente activos y Menores de 15 años de edad.]

Tabulado 2. CONDICIÓN DE ACTIVIDAD DE LA POBLACIÓN DE 15 AÑOS Y MÁS DE EDAD EN LA REPÚBLICA, POR ÁREA Y SEXO, SEGÚN PROVINCIA: ENCUESTA DE HOGARES, AGOSTO 2002(P).

Módulo Factor de Expansión: Se desarrollaron programas para manejar los factores de expansión; aplicable a los tres rangos de edades, a saber: menos y más 10 años de edad; de 10 a 14 años de edad; menos y más 15 años de edad. Aunque el más usado es el último.

Una vez que la base de datos queda lista para ser explotada, se generan (en aproximadamente 6 semanas) un promedio de 41 cuadros estadísticos, que conforman el Boletín de la Encuesta de Hogares. La publicación del Boletín en sí, toma de 3 a 5 meses.

Para el desarrollo óptimo de los Sistemas antes mencionados, se utilizó metodología de autoverificación. Además se utilizó el software Visual Foxpro (versión 5.0).

La decisión de utilizar los softwares de EXCEL y Visual Foxpro se debió a la capacidad que poseen de manejar gran cantidad de datos y de interactuar entre ellos.

Según la necesidad del usuario final, se crean diccionarios de datos utilizando el paquete de IMPS. El paquete SPSS es utilizado por usuarios locales para el análisis de los datos.

Al final del presente informe se adjunta diagrama general de los Sistemas antes mencionados.

Equipo (“hardware”) y programas (“software”):

Se utilizaron 8 PC (captura de datos):

- Micro-procesador Intel Pentium I de 166 MHZ
- RAM de 32 Mega

Se utilizó 1 PC (desarrollo de sistemas):

- Micro-procesador Intel Pentium III de 1.0 GHZ
- RAM de 128 Mega (Memoria Alta)

Se utilizó un Servidor:

- Micro-procesador Intel Pentium II de 450 MHZ (con 2 procesadores)
- RAM de 256 Mega Byte (Memoria Alta)
- El cableado de la RED era Estructurado, con un ancho de banda de 128 MB por segundo.

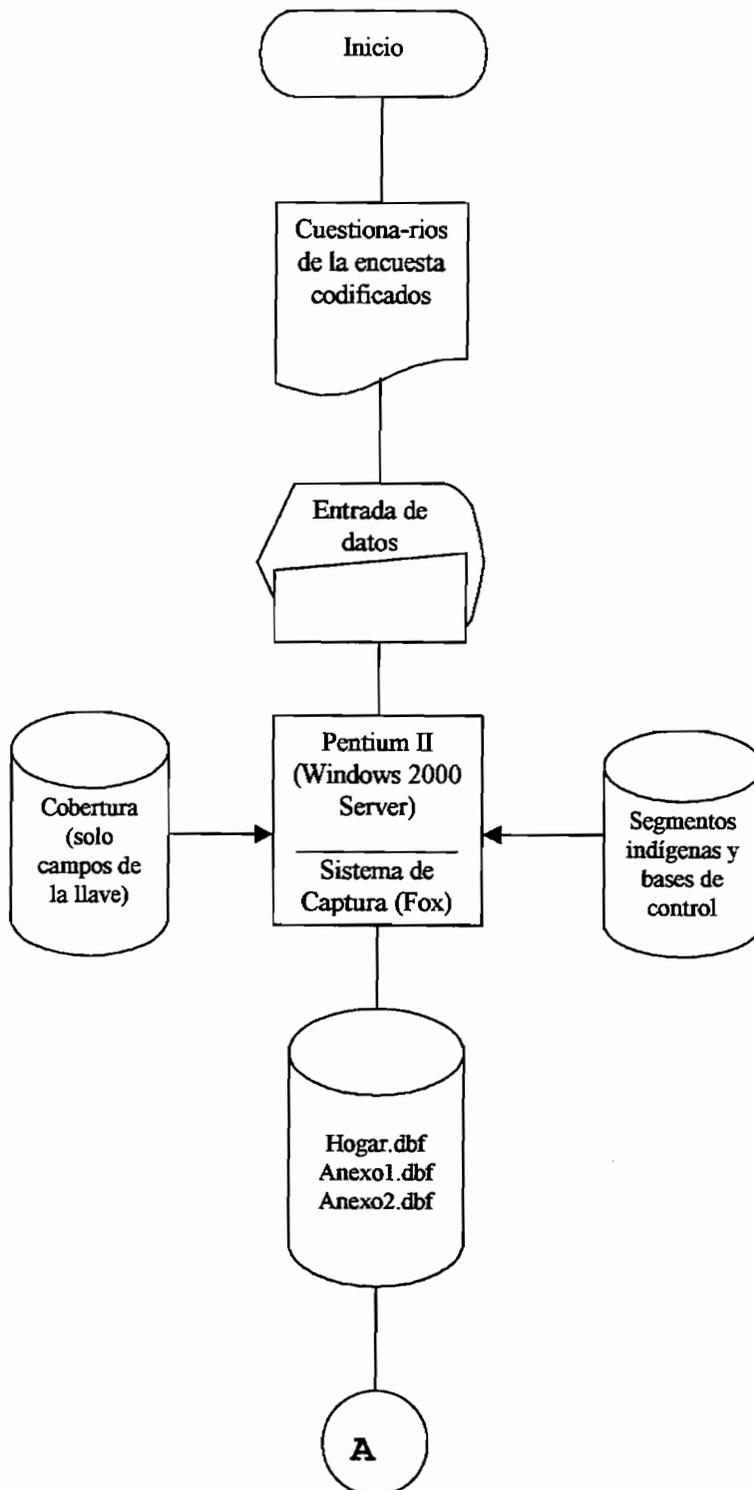
Para Internet se utiliza una línea Dedicada.

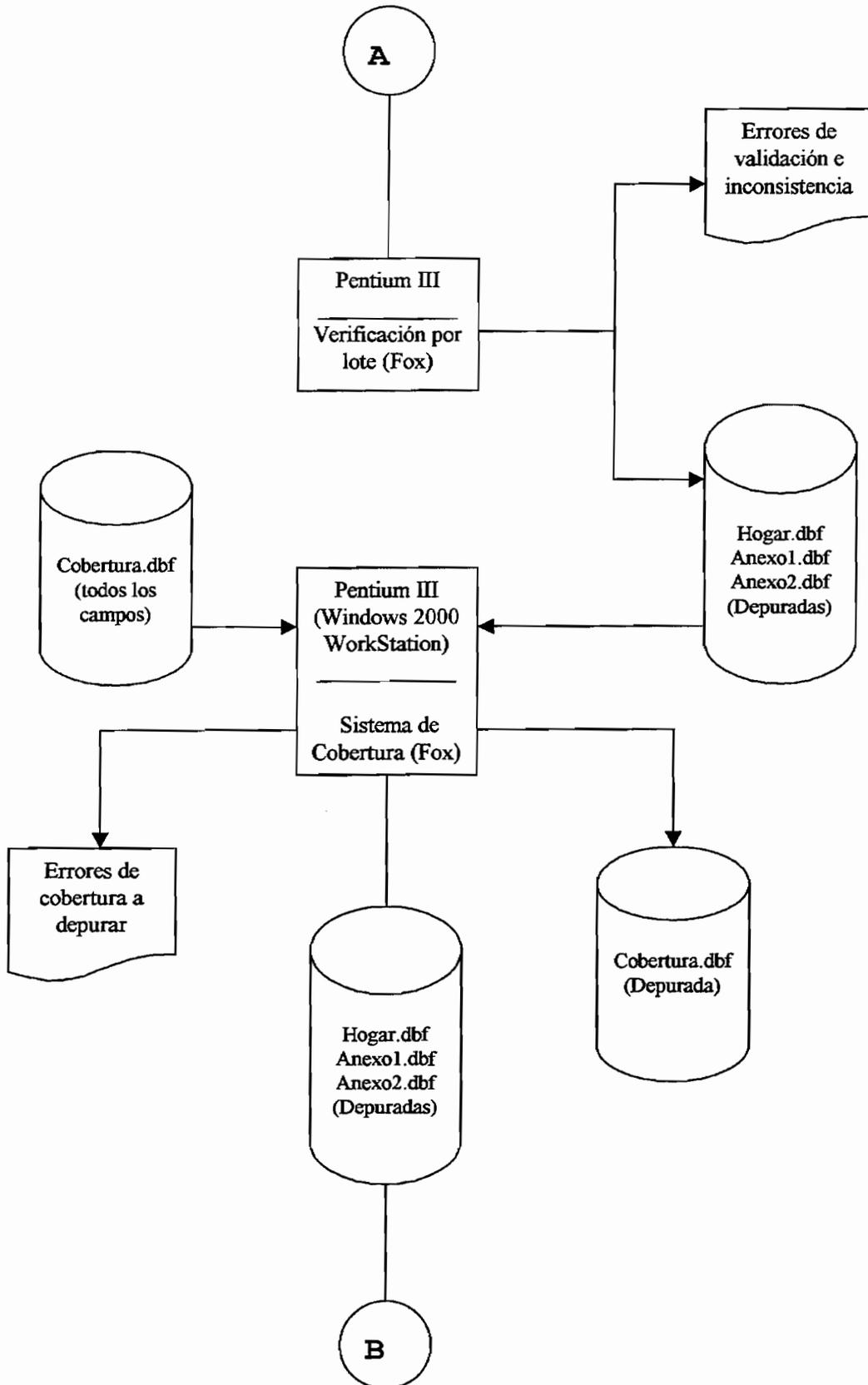
Sistemas operativos utilizados fueron Windows 2000 Server y Windows 2000 Workstation.

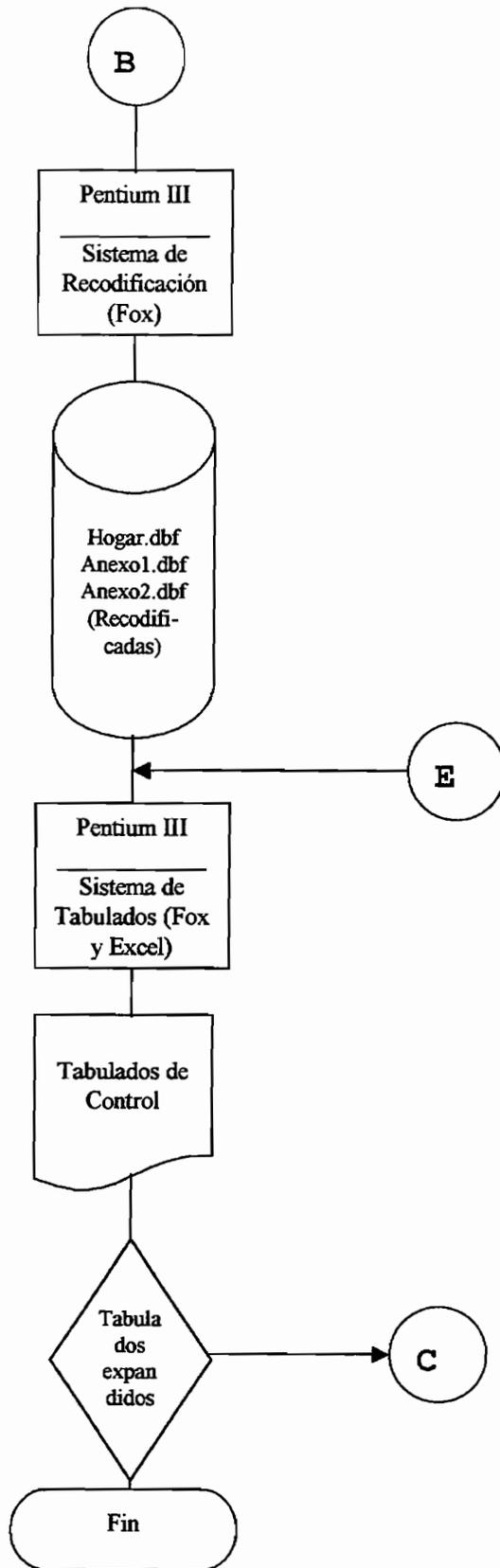
Para comprimir archivos se utiliza el software WinZip, versión 8.0

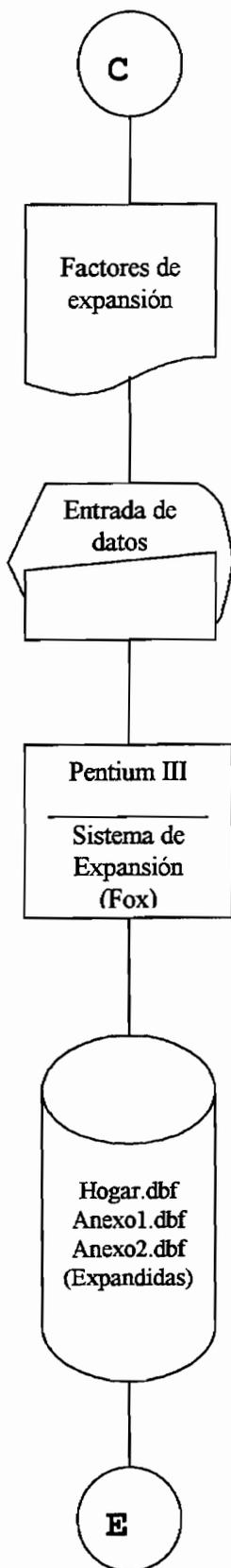
Se cuenta con un grabador de CD: CD Writer Plus, Serie 9200 de 8x. Marca Hewlett Packard.

DIAGRAMA DE SISTEMA ENCUESTA DE HOGARES-AGOSTO 2002









**ENCUESTA INTEGRADA DE HOGARES 2000/1
PROCESAMIENTO DE DATOS**

OSCAR ESTEBAN CORREA

DIRECCIÓN GENERAL DE ESTADÍSTICAS, ENCUESTAS Y CENSOS (DGEEC)

PARAGUAY

ÍNDICE

	<u>Página</u>
I. Encuesta Integrada de Hogares 2000/1.....	189
Antecedentes	189
Objetivos	189
Características	190
Cuestionarios.....	191
Cuestionario de hogares	192
Cambios en el cuestionario	194
II. Estrategias para la Captura de Datos	194
Codificación de las preguntas	195
III. Procedimientos para evaluar la consistencia de la información.....	197
IV. Diseños Computacionales de las Bases de Datos.....	198
V. Métodos para la Imputación de Datos	201
Tratamiento de la no-respuesta total y parcial	201
Criterios que se aplican para la imputación de datos faltantes	201
VI. Integración de Variables: Gasto, Consumo e Ingreso. Desocupados, Subempleados, etc. .	202
VII. Arquitectura de las Bases de Datos	203
VIII. Generación de Tabulados.	206
IX. Políticas Institucionales de Acceso a las Bases de Datos.....	206
Anexo 1	208
Plan de consistencia	208
Anexo 2	211
Variables que intervienen en la definición de Actividad Económica	211

I. Encuesta Integrada de Hogares 2000/1

Antecedentes

La Encuesta de Hogares tiene como propósito generar indicadores que permitan conocer la evolución del bienestar y las condiciones de vida de los hogares particulares y constituyen una herramienta poderosa para la toma de decisiones. En nuestro país la Encuesta Integrada de Hogares se realizó por primera vez entre los meses de agosto de 1997 a julio de 1998 donde la población investigada la constituye un grupo de personas residentes en viviendas particulares, ésta cubre las áreas Urbanas y Rurales de todo el país, excluyendo los departamentos de Alto Paraguay y Boquerón. Además fue integrada para proveer resultados para Asunción, Dpto. Central Urbano Rural, Resto del País Urbano Rural y los departamentos Alto Paraná, Itapúa, San Pedro y Caaguazú.

El cuestionario utilizado contaba de 9 secciones bien diferenciadas que relevaba información acerca de Características de la Población, Vivienda y Disponibilidad de Bienes Duraderos, Educación, Salud, Salud Infantil y Antropometría, Empleo e Ingreso Laboral, Actividades Independientes del Hogar, Actividades Agropecuarias Forestales y Derivados y Gastos del Hogar.

El tamaño de la muestra a nivel nacional fue de 5.000 viviendas y fue distribuida en estratos, considerando las zonas Urbanas y Rurales del Total del País de acuerdo al Censo de 1992 y además, con una aproximación cercana a la asignación de Neyman. En la primera etapa se seleccionaron en total 500 segmentos para los cinco estratos y en la segunda etapa el tamaño de la muestra fue variable, dependiendo de la cantidad de viviendas por conglomerados para cada uno de los estratos.

Las estimaciones vienen afectadas por dos tipos de errores: los errores de muestreo y los errores ajenos al muestreo.

Objetivos

Una Encuesta de *Estudios sobre la Medición de las Condiciones de Vida (LSMS)* tiene como propósito el de explorar formas de perfeccionar, tanto la calidad, como el tipo de datos sobre hogares recogidos por las oficinas estadísticas y tiene como meta promover el uso creciente de la información acerca de hogares. La EIH utilizando esta metodología y adaptada a nuestro contexto, proporciona información adecuada para la planificación, seguimiento y análisis como fundamento para la toma de decisiones en materia de políticas y programas de apoyo social con respecto al impacto que ejercen sobre las condiciones de vida de los hogares, especialmente los más pobres

Esta Encuesta contó con el apoyo técnico y financiero del Banco Interamericano de Desarrollo (BID) y el Banco Mundial (BM) en el marco del proyecto *Mejoramiento de las Encuestas de Condiciones de Vida (MECOVI)*.

El Programa MECOVI es un Programa Regional para América Latina y el Caribe que tiene como objetivos principales:

- Fortalecer los sistemas de encuestas de hogares en los países de la región, con el fin de generar información amplia y completa sobre las condiciones de vida y pobreza de la población con el propósito de disponer de información útil para la evaluación y formulación de políticas en el área social.
- Capacitar a técnicos nacionales para mantener y operar de modo eficiente el sistema de encuestas de hogares.
- Asistir y capacitar al personal local en el análisis y uso de la información proveniente de las encuestas para el diseño, elaboración, seguimiento y evaluación del impacto de las políticas orientadas a reducir la pobreza y la desigualdad.

Características

Por tratarse de una Encuesta Integrada de Hogares sus características son bien definidas: el **Ámbito Geográfico**, **Ámbito Poblacional**, **Periodos del Levantamiento de la Información**, los **Periodos de Referencias**.

A continuación se hace una rápida narración de estas características.

Ámbito Geográfico: Cubre todas las áreas Urbanas y Rurales de todo el país.

Ámbito Poblacional: La población investigada es la constituida por el grupo de personas residentes en viviendas particulares, se excluye la población residente en viviendas colectivas.

Periodos de Levantamiento: La Encuesta Integrada de Hogares 2000/2001 se llevó a cabo entre los meses de septiembre de 2000 a agosto de 2001.

Periodos de Referencia:

- ✓ Día de la Encuesta
- ✓ Último año
- ✓ Últimos 5 años
- ✓ Mes anterior
- ✓ Presente año
- ✓ Últimos 30 días
- ✓ Últimos 90 días
- ✓ Últimos 7 días
- ✓ Últimos 7 meses
- ✓ Último mes

Niveles de Estimación:

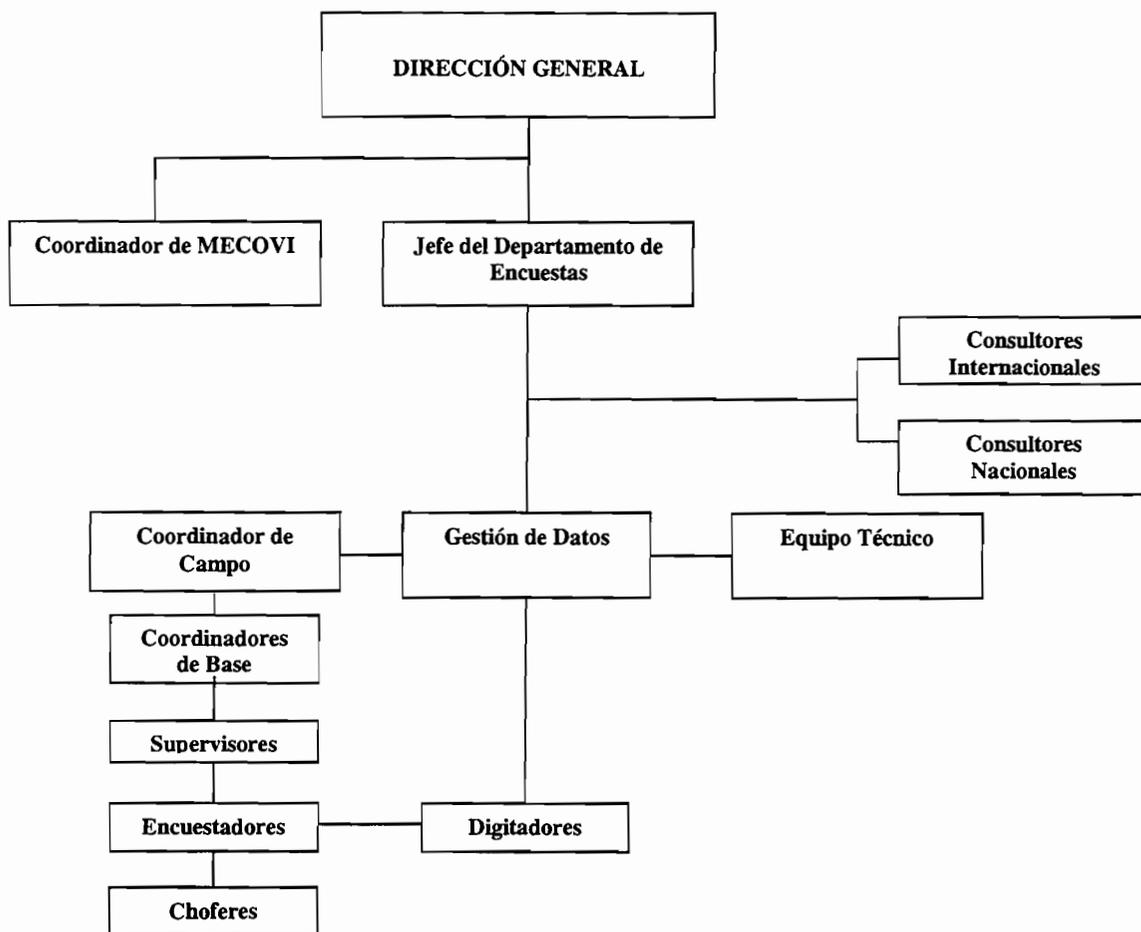
- ✓ Total País
- ✓ Total Urbana
- ✓ Total Rural
- ✓ Departamento (excepto Alto Paraguay y Boquerón)

Tipo de Muestreo: bietático y probabilístico, en la primera etapa se eligen por conglomerados de viviendas con probabilidades iguales a su tamaño, en términos de números de viviendas; en la segunda etapa, mediante submuestreo, se selecciona un número fijo de viviendas sin reposición y con equiprobabilidades.

Tamaño de la Muestra: 8.960 viviendas.

Recolección y Tratamiento de los Datos: Las viviendas de la muestra son visitadas por Encuestadores, quienes recogen los datos de la vivienda y de cada persona que habitan en las mismas, en un cuestionario destinado para cada hogar. Los datos se ingresan inmediatamente después de la entrevista.

Organización



Cuestionarios

Con el objeto de recoger información consistente, la EIH utiliza tres tipos de cuestionarios:

- el cuestionario de hogares
- el cuestionario comunitario
- el cuestionario de precios

Cuestionario de hogares

Éste recoge datos de los miembros del hogar respecto a varios aspectos de su bienestar y condiciones de vida. Contaba inicialmente de 13 secciones con un total de 66 páginas que a continuación se detallan.

Secciones / Módulos	Parte	Descripción	Población Investigada	Periodo de Referencia
CARACTERÍSTICAS DE LA POBLACIÓN		Ficha del Hogar Migración	Todos los miembros del hogar	- Día de la Encuesta - 1 y 5 años anterior a la encuesta.
CARACTERÍSTICAS DE LA VIVIENDA	A	Vivienda	Vivienda	- Día de la Encuesta - Mes anterior - El último año
	B	Inventario de Bienes Duraderos	Vivienda	- Día de la Encuesta
EDUCACIÓN		Escolaridad Gastos Curso No Formal	Personas de 5 años y más	- Día de la Encuesta - El presente año - Los últimos 30 días
SALUD	A	Salud de todos los Miembros	Todos los miembros del hogar	- Los últimos 90 días
	B	Mujeres	Mujeres de 15 a 49 años	- Los últimos 5 años - Día de la Encuesta
SALUD INFANTIL Y ANTROPOMETRÍA	A	Salud Infantil	Menores de 5 años	- Los últimos 90 días - Día de la Encuesta
	B	Antropometría		
EMPLEO E INGRESO LABORAL	A	Resumen de las Actividades Económicas de los Miembros del Hogar	Personas de 10 años y más	- Los últimos 7 días
	B	Trabajo Principal Agropecuario y No Agropecuario		- Los últimos 7 días - Los últimos 12 meses
	C	Trabajo Secundario		- Los últimos 7 días
	D	Subempleo y Búsqueda de Trabajo Suplementario		- Los últimos 7 días
	E	Ingresos del Trabajo y Otras Fuentes		
	F	Ingresos Extraordinarios		- Los últimos 12 meses
ACTIVIDAD INDEPENDIENTE DEL HOGAR	A	Información de la Empresa	Empresas Independiente del Hogar	- Los últimos 12 meses
	B	Actividades, Gastos e Ingresos de la Empresa durante el Último Mes		- El último mes
		Actividades, Gastos e		

	C	Ingresos de la Empresa durante los Últimos 12 Meses		- Los últimos 12 meses
ACTIVIDAD AGROPECUARIA, FORESTAL Y DERIVADOS	A	Información sobre Lotes	Actividades Agropecuarias del Hogar	- Los últimos 12 meses
	B	Maquinarias E Implementos Agrícolas		- Día de la Encuesta
	C	Animales y Producción Pecuaria		- Los últimos 90 días - Día de la Encuesta
	D	Procesamientos de Productos Agropecuarios y Forestales		- Los últimos 12 meses
	E	Insumos		- Los últimos 30 días
	F	Alquiler de Recursos		- Los últimos 12 meses
	G	Servicios		- Los últimos 12 meses
	H			- Los últimos 12 meses
GASTOS DEL HOGAR	A	Gastos Alimentarios del Hogar	Hogar	- Los últimos 30 días - Los últimos 7 días
	B	Gastos Alimentarios Fuera del Hogar	Personas de 10 años y más	- Los últimos 7 días
	C	Gastos No Alimentarios	Hogar	- Los últimos 30 días
	D	Gastos No Alimentarios Referidos a los Últimos 12 Meses	Hogar	- Los últimos 12 meses
	E	Gastos por Transferencias A Terceros	Personas de 10 años y más	- Los últimos 12 meses
SEGURIDAD Y VIOLENCIA		Extra Familiar	Hogar	- Los últimos 5 años
SITUACIONES ADVERSAS		Situaciones Adversas del Hogar	Hogar	- Los últimos 12 meses
OTRAS TRANSACCIONES DEL HOGAR		Transacciones del Hogar	Hogar	- Los últimos 12 meses
PARTICIPACIÓN CIUDADANA		Participación Ciudadana	Personas de 10 años y más	- Día de la Encuesta

Cambios en el cuestionario

Debido al recorte presupuestario a partir del periodo 19 (56 en total) de la Encuesta, fueron excluidas del cuestionario las secciones:

9. Gastos del Hogar, con excepción del rubro de Yodación de la sal página 55 y Parte E, página 62, que corresponde a Gastos por Transferencias a Terceros.
10. Seguridad y Violencia,
11. Situaciones Adversas,
12. Otras Transacciones del Hogar y
13. Participación Ciudadana.

En el periodo 29 fue incluida la Sección 14. Evaluación de Servicios por Empresas del Estado en provisión a los Servicios Básicos.

Sección / Módulo	Parte	Descripción	Población Investigada	Periodo de Referencia
EVALUACIÓN DE SERVICIOS		Servicios Básicos	Hogar	Día de la Encuesta

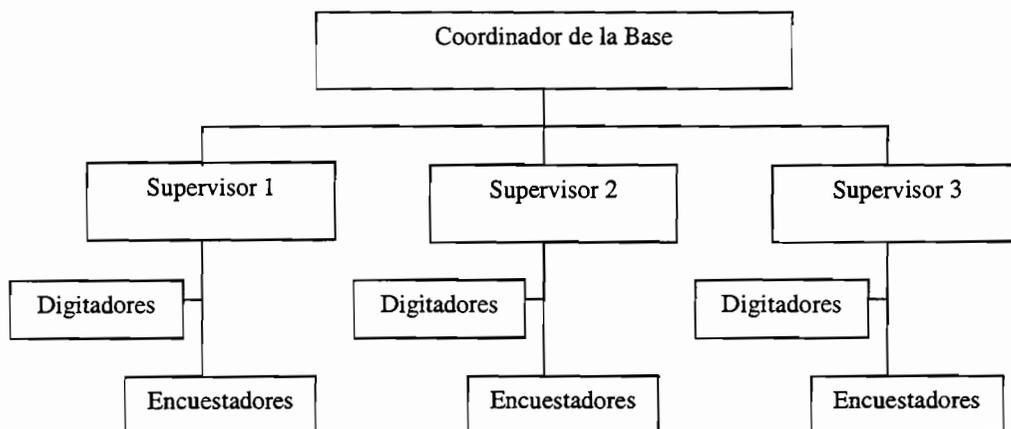
En total el cuestionario está compuesto de 587 preguntas aproximadamente y su llenado se estima de 4 a 5 horas en dos entrevistas.

II. Estrategias para la Captura de Datos

La Encuesta de Hogares hace uso de un conjunto de procedimientos destinados a minimizar los errores y retrasos en la recolección y procesamiento de datos; uno de ellos es el uso de los adelantos en la tecnología informática aplicada a los trabajos de campo. A partir de la Encuesta Integrada de Hogares del año 1997, la entrada de datos se realiza en forma inmediata a la recolección en terreno.

Como la Dirección de Estadística, Encuestas y Censos (DGEEC) no cuenta con oficinas regionales, se recurre a instituciones que permiten establecer una oficina operativa temporal.

La conformación de los grupos de trabajo de campo estaba organizada de la siguiente manera:



Bases	Equipos	Supervisores	Encuestadores	Digitadores	Choferes
Encarnación	3	3	9	3	3
Choré	2	2	6	2	2
Oficina Central	5	5	15	5	5
Coronel Oviedo	3	3	9	3	3
Hernandarias	3	3	9	3	3

Un digitador acompaña a cada equipo en el trabajo de campo. Cuando el trabajo de campo no requiere pernocte, el digitador ingresa los datos en la oficina, en caso contrario se realiza en el lugar de pernocte.

Codificación de las preguntas

En la mayor parte del cuestionario las respuestas potenciales de las preguntas están precodificadas, de forma que el entrevistador tenga que marcar con una 'X' la opción o registrar el código de la respuesta en el cuestionario. La mayoría de los códigos están escritos en el recuadro de la respectiva pregunta.

PARTE A: VIVIENDA

1 TIPO

Casa 1

Rancho 2

Dpto. o piso 3

Pieza o inquilinato 4

Vivienda improvisada 5

Otro (especificar) 6

PARA TODAS LAS PERSONAS

3	4
¿Qué relación de parentesco tiene ...[NOMBRE]... con el Jefe del Hogar?	¿ES MIEMBRO DEL HOGAR ...[NOMBRE]...?
Jefe/a 1 Esposa/a/compañera/a 2 Hijo/a 3 Otro pariente 4 No pariente 5 Empleado doméstico 6	SI 1 NO 6 (Fin entrevista)

1	
2	

Cuando la lista de códigos se presenta muy extensa como para insertarlo en el recuadro o ésta se aplique a varias preguntas se ubica en el reverso de la página anterior, de tal forma que quede visible cuando el entrevistador esté trabajando en la página referida.

Pocas preguntas requieren de codificación manual (códigos geográficos, rama de actividad, actividad económica).

<div style="border: 1px solid black; width: 30px; margin: 0 auto; padding: 2px;">1</div> <p>¿Podría informarme sobre el TRABAJO PRINCIPAL que hizo durante los últimos 7 días o la ocupación que tiene actualmente?</p> <p>¿Qué hace en este trabajo ?</p> <p>Por ejemplo: - secretaria - recepcionista - zapatero - mecánico chapista - gerente</p>	<div style="border: 1px solid black; width: 30px; margin: 0 auto; padding: 2px;">2</div> <p>¿A qué se dedica el establecimiento o empresa en el que trabaja en su OCUPACION PRINCIPAL ?</p> <p>Por ejemplo: - fábrica de aceite - confección de ropas - oficina gubernamental</p>		
DESCRIPCION DE LA OCUPACION	CODIGO (Dejar en blanco)	ACTIVIDAD ECONOMICA	CODIGO (Dejar en blanco)
1			
2			

Los digitadores son capacitados para realizar estas codificaciones y cuentan con los siguientes manuales:

- Manual de códigos geográficos (Países, Departamentos y Distritos)
- Manual de códigos de Ocupación
- Manual de códigos de Actividad Económica

Para facilitar esta tarea, además de los manuales de codificación proporcionados, los digitadores son asistidos por un programa diseñado para el efecto. El programa accede a la base de datos que contiene los manuales de codificación y realiza la búsqueda de la cadena o palabra solicitada, proporcionando todos los códigos cuya descripción contengan la cadena. Con esta ayuda el digitador dispone de una gama de alternativas y selecciona aquella que considera apropiada.

En los casos donde la precodificación no incluya todas las posibles respuestas, se agrega un código para "Otro (especifique)". Finalizado el trabajo de campo, este código es listado con su respectiva especificación para ser analizados y recodificados si amerita hacerlo.

24

Además de lo que aprendió en el curso, ¿cuál fue el principal beneficio que obtuvo...[NOMBRE]... de la capacitación que recibió?

Conseguir trabajo 1
 Crear empresa propia 2
 Obtener ingresos adicionales 3
 Mejorar su ingreso 4
 Promocionarse en el trabajo 5
 Mejorar su empresa 6
 Otro (especificar) 7
 Ninguno 8

▶ SGTE. SECCION

	1
	2

III. Procedimientos para Evaluar la Consistencia de la Información

El programa de entrada de datos fue diseñado de tal forma que cada pantalla está en función a una página y se acceden a las mismas a través de menú. Si en la página existe diferentes estructuras de preguntas, se generan tantas pantallas de entrada de datos como sean necesarias. Cada pantalla genera un tipo de registro.

Para la construcción del nombre de las variables se estableció colocar al número de pregunta un prefijo que identifique la sección a la que pertenece.

	Prefijo
Sección 1, Características de la Población	: P
Sección 2, Características de la Vivienda	: V
Sección 3, Educación	: ED
Sección 4, Salud	: S
Sección 5, Salud Infantil y Antropometría	: SI

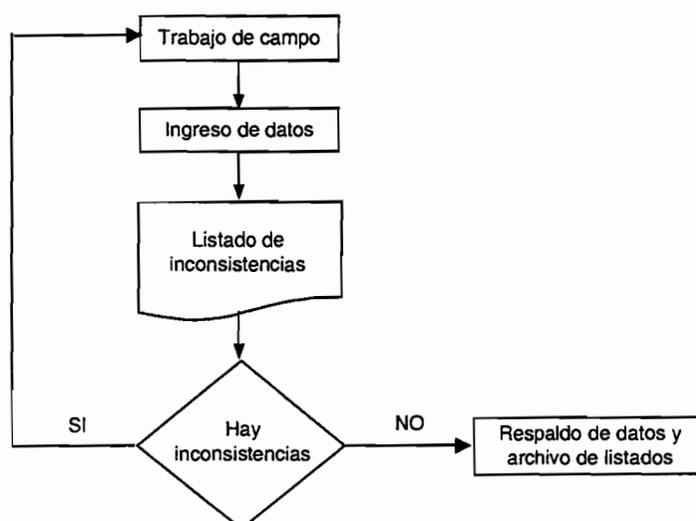
Así la denominación de la pregunta 1 de la Sección 1 es P01

El programa de entrada de datos realiza controles de flujo, rango y consistencia de los datos. Estos controles pueden ser a nivel de cada registro o entre registros, de la misma Sección o Secciones diferentes.

Los mensajes emitidos fueron diseñados lo más claro posible, de tal forma que el digitador y el supervisor requieran una mínima asistencia de los técnicos de la oficina central.

El digitador puede corregir solamente los errores u omisiones cometidos por él durante el proceso de carga, los demás errores son identificados y señalados en el cuestionario para que sean verificados por el supervisor del equipo y corregidos e ingresados inmediatamente en el terreno. Este proceso se repite hasta que el programa no registre inconsistencias.

Diagrama del proceso de control de inconsistencias



Los mensajes de error que no requieran modificación de los datos ingresados, deben ser justificados fehacientemente por el supervisor en el listado de inconsistencias.

El programa de entrada de datos dispone de varias tablas de control:

- Tabla de control de los datos referentes a cada Unidad Primaria de Muestreo (Código Geográfico, Equipo, Período y Viviendas Seleccionadas).
- Tabla de correspondencias posibles entre la ocupación, la actividad económica, la categoría ocupacional, el nivel educativo y el tamaño de empresa. De haber inconsistencia, emite un mensaje para que el digitador verifique la codificación y/o consulte con algún técnico de la oficina central. Si es necesario, la tabla de correspondencia se actualiza y se instala nuevamente en todos los computadores.
- Tabla de precios (valor mínimo y máximo) de productos alimenticios por unidad de medida
- Tabla de precios (valor mínimo y máximo) de bienes duraderos

Un 10% de las viviendas seleccionadas son revisadas por el supervisor. Los datos de estos cuestionarios son ingresados y cotejados por el programa con los datos de la entrevista. De encontrar diferencia se procede a verificar en el terreno la causa de la misma. Este procedimiento permite al supervisor detectar a tiempo, errores conceptuales o sistemáticos que pueda cometer el encuestador en el terreno.

Otro procedimiento para el control de calidad de los datos es el seguimiento que el equipo de técnicos de la oficina central realiza periódicamente en las bases operativas. Durante estas visitas se supervisan todos los cuestionarios digitados (entrevistas y reentrevistas) y sus respectivos listados de inconsistencias, se verifican las codificaciones y las justificaciones de las inconsistencias. En caso de duda, el coordinador de campo verifica nuevamente los datos en el terreno. En estas visitas se recogen los datos de los cuestionarios ya ingresados que no requieran de corrección o verificación. Con estos datos se generan los archivos para análisis y tabulación en la oficina central, que permiten controlar posibles inconsistencias.

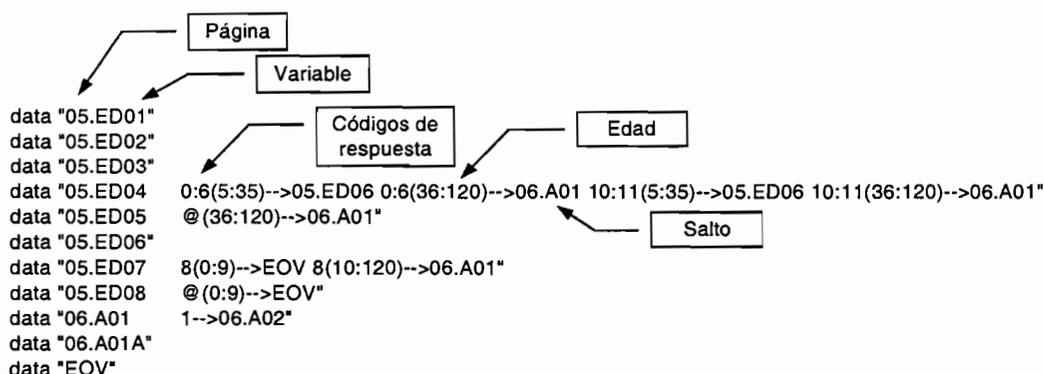
Si bien los programas de consistencias son evaluados con las pruebas pilotos, no siempre contemplan todos los casos posibles de la realidad. Por ello, en la medida en que se detectan casos no cubiertos, se generan programas de seguimiento para los archivos digitados y se actualiza el programa de entrada y consistencia de datos.

Durante la captura de datos no se generan archivos temáticos. Tampoco se agregan datos ni se forman variables.

IV. Diseños Computacionales de las Bases de Datos

Para la Encuesta Integrada del año 1997 a sugerencia del consultor chileno Juan Muñoz pusimos a prueba un programa de entrada y consistencia de datos desarrollado por él. Este programa reúne ciertas características que lo hacen atractivas para su empleo en las Encuestas con digitación en terreno:

- Los controles del flujo de la información a nivel "Cuestionario" son más fáciles de aplicar, considerando que la Encuesta Integrada es multitemática, y ciertos módulos, en particular el de las personas, son aplicables a diferentes grupos de edad e inclusive sexo (Fecundidad: mujeres de 15 a 49 años, Salud infantil: menores de 5 años). Este programa requiere realizar simplemente una tabla en la que se enumeran ordenadamente por página o tipo de registro las variables, dependiendo del código de la respuesta, el sexo o la edad, se le indica a qué página y pregunta debe saltar. Procedimiento que nos ahorra tiempo y facilita la realización de estos controles.



Ejemplo del mensaje emitido cuando se produce un error de salto:

---- mensaje 5 (Inconsistencia: P30.SA) ----

ERROR, la persona N° 01 (Tomasa Noel de Martínez, mujer de 62 años) no respondió la pregunta C01D de la página 30.

El flujo de la información también puede controlarse en cada pantalla. Ejemplo de una porción de la definición de variables y saltos en la pantalla de la Página 5.

```

V18, "18 Sit. legal vivienda ", 06,14,44,01,QLN
1, "Propia ",-->V21
2, "Pagando cuotas ",-->V20
3, "En condominio ",-->V21
4, "Arrendatario ",-->V19
5, "Ocupante de hecho ",-->V23
6, "Cedida ",-->V23
7, "Otro "
EOC

V18E," Especifique ", 07,14,45,15,TYP,-->V23
V19, "19 Guaraníes alquiler ", 09,14,44,08,QNI
00000000,00000000,-->V25
00020000,03000000,-->V25
EOC
  
```

- Los mensajes de inconsistencias emitidos pueden redactarse de tal forma que el digitador o supervisor dispongan de toda la información requerida para la toma de decisión durante el trabajo de campo. Los listados de inconsistencias generados pueden ser analizados en pantalla o formato impreso.

---- mensaje 5 (Inconsistencia: P26.10) ----

ERROR, La persona N° 04 (Filomena Rivero Godoy, mujer de 56 años) se registró en P26.12 como empleado doméstico y en la P29.26 que vive allí, sin embargo no se registró como empleado doméstico en P02.03.

---- mensaje 6 (Inconsistencia: P44G.01) ----

ERROR, En la página 44G falta la línea 8.

---- mensaje 7 (Advertencia: P36.01) ----

VER: Se registra la presencia de trabajadores independientes en P36.01, a pesar de que ningún miembro del hogar declaró realizar actividades independientes en los últimos 7 días como Patrón o Cta. Propista en la P26.12 o P30.06.

---- mensaje 8 (Advertencia: P42.13) ----

VER: No se ha registrado la cosecha del cultivo tipo 107 (maní), a pesar de que se ha adquirido insumos para los mismos en la página 43F.

- Dispone de un módulo de control para las mediciones antropométricas que recibe como parámetros la edad, talla y peso.

---- mensaje 4 (Inconsistencia: P21.01) ----

ERROR, la persona N° 04 (Shirley Aide Romero, niña de un año) es muy pequeña para su edad.

---- mensaje 6 (Inconsistencia: P21.01) ----

ERROR, la persona N° 03 (Gisella Valiente, niña de 4 años) es muy pesada para su edad.

---- mensaje 8 (Advertencia: P21.01) ----

VER, la persona N° 03 (Gisella Valiente, niña de 4 años) parece muy alta para su edad.

- Facilita la inserción de registros, en este programa se administra por medio de índices. Por ejemplo se puede ingresar primero la persona de la línea 5 y posteriormente la de la línea 1. Internamente él los ordena. Esto facilita el proceso de carga; en la medida que se disponen de datos en el cuestionario los mismos son digitados en forma inmediata, de esta forma el digitador no necesita disponer del cuestionario con llenado completo.

En el IMPS esto requiere posicionarse en uno de los registros cargados y direccionar si el registro nuevo se insertará antes o después, proceso que requiere de mayor atención por parte del digitador.

- El chequeo entre los datos de la entrevista y la reentrevista requiere únicamente disponer de una tabla en la que se enumeran ordenadamente las variables por páginas y se indica cual que variable debe comparar. El resultado de la comparación puede ser analizado en pantalla o formato impreso.

Con anterioridad este proceso se realizaba a través de un programa diseñado en lenguaje COBOL.

- El programa permite definir en el diseño de pantalla el tipo de variable de acuerdo a la consistencia necesaria:
 - Cronológicas: Para las variables tipo fecha.
 - Tipográficas: Para los textos.
 - Cuantitativas: Para la variables numéricas sin rango definido.
 - Cuantitativas por intervalo: Para las variables numéricas con rango definido.

Horas que trabaja al día: 00,15:00,59
 01,00:01,59
 02,00:02,59

 23,00:23,59
 24,00:24,00

- Cualitativas: Para la variables numéricas con categorías. En el momento de digitar el código se despliega en pantalla la categoría.

- 1, "Empleado público"
- 2, "Empleado privado"
-
- 7, "Trabajador familiar no remunerado"
- 8, "Empleado doméstico"

Para la elaboración del programa de consistencias, personal técnico del Departamento de Encuestas y del Departamento de Informática redactan la documentación guía para las mismas.

A modo de ejemplo ver ANEXO 1

V. Métodos para la Imputación de Datos

Tratamiento de la no-respuesta total y parcial

Para disminuir el efecto de posibles sesgos por la presencia de la no-respuesta, en el marco de las encuestas de hogares ejecutadas por la DGEEC, se procede de manera preventiva, es decir, en el momento que se determina el tamaño de la muestra, se prevé la posibilidad de que se produzcan rechazos:

- totales, al inicio de la entrevista (después de agotar todos los mecanismos preestablecidos para que no se produzca) o,
- parciales (cuando la información captada no alcance a cubrir el módulo de empleo).

Criterios que se aplican para la imputación de datos faltantes

Si el rechazo es parcial, en alguna pregunta específica, el tratamiento es diferente. Se imputan valores sólo a las preguntas sin respuestas que participan en la construcción de los ingresos y gastos agregados, manteniendo las variables originales en las bases de datos, es decir, creando variables nuevas con los valores imputados. En el resto de las preguntas en general, no se hace imputación alguna, debido a la inexistencia o escasa frecuencia de los casos con no-respuesta. Cuando el supervisor agota las posibilidades de obtener el dato, se registra en la pregunta "N/D". El digitador en estos casos pulsa la tecla 'F10' que llena el campo con asteriscos (*), tantos como longitud tenga el campo. En el proceso de unión de registros, cada asterisco es reemplazado por '9'

El procedimiento de imputación a las preguntas que participan en la construcción de los ingresos y gastos agregados es el siguiente:

- Se determina el número de casos con no-respuesta para cada una de las preguntas analizadas con el propósito de aplicar luego el método más pertinente.
- Cuando es *ínfimo* el número de casos con no-respuesta no se hace imputación alguna. Por ejemplo, las preguntas de gasto referidas a alimentos, no alimentos y salud en 1997/98 que no tuvieron respuesta no alcanzó al 0,1 % de los casos.
- Cuando es *pequeño* el número de casos con no-respuesta se construyen matrices desagregadas con las medianas obtenidas sobre la base de los casos (expandidos) que tienen respuesta, teniendo en consideración variables independientes relevantes. Por ejemplo, a las preguntas de gasto en agua y combustible sin respuesta (0,8 % de los casos) se imputa la mediana obtenida con los valores mayores a cero según el tipo de abastecimiento y el dominio geográfico de residencia.
- Cuando es *importante* el número de no-respuestas se construyen regresiones específicas teniendo en consideración variables independientes relevantes. Por ejemplo, a la pregunta sobre "cuánto estima que le pagaría en concepto de por su vivienda" sin respuesta (generalmente entre 10 % y 12 % de los

casos) se imputa el valor proyectado de un modelo semilogarítmico del “alquiler mensual ...” con las siguientes variables independientes: tipo de vivienda, número de dormitorios, material de las paredes, piso y techo; tipo de abastecimiento del agua, disponibilidad de corriente eléctrica, teléfono, pieza para bañarse, ducha y agua caliente; tipo de servicio sanitario, disponibilidad de pieza para cocinar, tipo de propiedad de la vivienda y disponibilidad de piezas para destinarlo a actividades económicas.

Cuando se concluye el proceso de imputación, se comparan cada una de las variables originales con aquellas que contienen los valores imputados, en términos de la media, mediana y desviación estándar a nivel de cada dominio geográfico (es decir, en términos de la variable que se utiliza para estratificar la muestra durante el diseño muestral). Según las pruebas realizadas, nunca las diferencias de tales indicadores han sido estadísticamente significativas.

El proceso de imputación de las variables es documentada en formato de sintaxis (de SPSS, IMPS, etc.) y son entregadas a solicitud de los interesados.

Las variables originales, es decir, las que contienen la información recogida durante el trabajo de campo, después de consistenciarlas y limpiarlas, no son modificadas durante el proceso de imputación. Por tanto, los usuarios reciben las bases con las variables originales y con variables generadas durante el proceso de imputación, con los diccionarios y documentación correspondientes.

VI. Integración de Variables: Gasto, Consumo e Ingreso. Desocupados, Subempleados, etc.

En el proceso de unión de registros se integran las siguientes variables

- En la tabla de viviendas (R01_00) se agregan las siguientes variables

Necesidades Básicas Insatisfechas (NBI):

NBI1	: NBI en Calidad de la Vivienda	1: Tiene	6: No tiene
NBI2	: NBI en Infraestructura Sanitaria	1: Tiene	6: No tiene
NBI3	: NBI en Acceso a la Educación	1: Tiene	6: No tiene
NBI4	: NBI en Capacidad de Subsistencia	1: Tiene	6: No tiene
NBI5	: NBI Cantidad de NBI por hogar	1: Tiene	6: No tiene
NBIA1	: NBI en al menos 1	1: Tiene	6: No tiene

Tipo de hogar (TIPOHOGA):

- 1 : Unipersonal
- 2 : Nuclear completa
- 3 : Nuclear incompleta
- 4 : Extendido
- 5 : Compuesto

- En la tabla de población (R02_00) se agregan las siguientes variables

Actividad Económica Desagregada (PEAD):

- 1 : Ocupados (Excluye Subocupación Visible e Invisible)
- 2 : Desocupados de 2ª. o más veces
- 3 : Inactivos (Excluye Desempleo Oculto)
- 4 : Subocupación visible
- 5 : Subocupación invisible
- 6 : Desempleo Oculto
- 7 : Desocupados de 1ª. Vez
- 0 : NA

Actividad Económica Agrupada (PEAA):

- 1 : Ocupados
- 2 : Desocupados
- 3 : Inactivos
- 0 : NA

Informal (INFO):

- 1 : Informal
- 0 : No informal

- Ingresos y Gastos agregados

Desde la Encuesta de 1997/98 los ingresos y/o gastos agregados de los hogares se presentan en archivos generados de manera independiente, los cuales se entregan a los usuarios con el resto de archivos originales. Estos últimos contienen además la información sobre la agregación y las líneas de pobreza, los quintiles y deciles para el análisis de la distribución de los ingresos.

Todas las variables que se expresan en términos monetarios se convierten a precios del período medio de la encuesta (ejemplo: febrero de 1998 para la EIH 1997/98, marzo del 2001 para la EIH 2000/01, etc.) utilizando el Índice de Precios al Consumidor proporcionado por el Banco Central del Paraguay.

Este proceso de normalización en el tiempo, es documentado en formato de sintaxis conjuntamente con el proceso de imputación. En los diccionarios de variables son fácilmente detectables las formas o unidades en que se encuentran las variables normalizadas.

Con el fin de que los datos de las Encuestas sean comparables, los períodos de referencias se mantienen constantes en las principales variables de estudio. En Actividad Económica: los últimos 7 días; Actividad Independiente Agropecuaria y No Agropecuaria: los últimos 12 meses.

VER ANEXO 2 (Diagrama para definir la variable Actividad Económica - PEAA)

VII. Arquitectura de las Bases de Datos

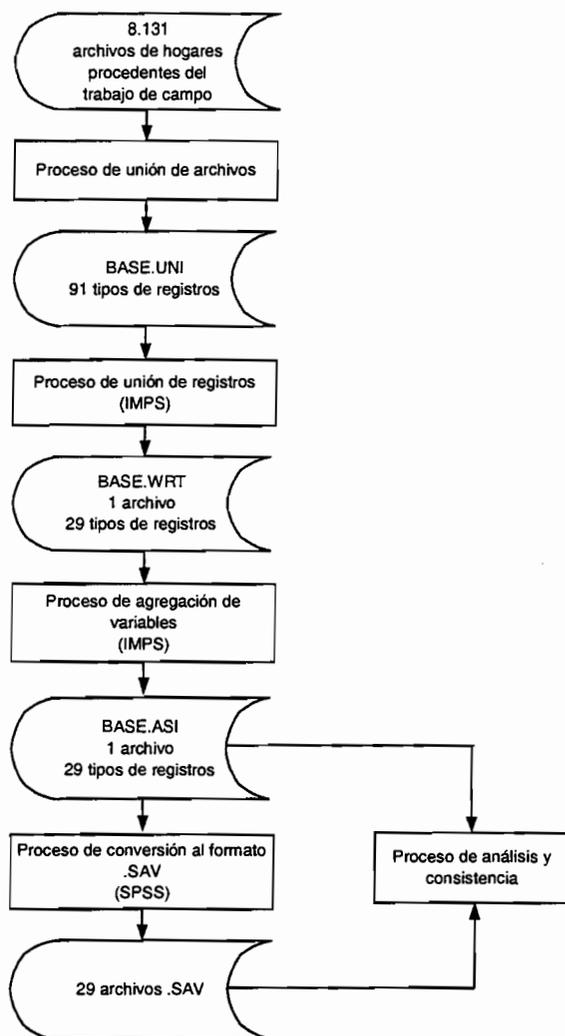
Tanto el paquete IMPS como el programa proporcionado por el Sr. Juan Muñoz generan archivos en formato ASCII, con identificadores de tipo de registro (1, Vivienda; 2, Población).

Cada hogar ingresado constituye un archivo, en el momento de cargar un hogar, el programa verifica con una tabla de correspondencia : el equipo, la clave geográfica, el número de vivienda seleccionada y el periodo de trabajo con la Unidad Primaria de Muestreo (UPM), con el fin de evitar la duplicación de archivos.

En el momento de recoger los archivos (informáticos y cuestionarios) de las bases operativas se verifica si cada cuestionario cuenta con su archivo correspondiente, en caso contrario se controla con la carpeta del supervisor la razón por la cual no fue ingresada (desocupada, abandonada, de veraneo, en construcción, en alquiler o venta, ocupantes ausentes, rechazo, otro).

Posterior a la recolección de archivos, éstos se unen, formando 1 sólo archivo en formato ASCII (BASE.UNI) con diferentes identificadores de registro (91 tipos de registros).

Diagrama de generación de archivos



El archivo unido (BASE.UNI) es procesado en IMPS para generar otro archivo en formato ASCII uniendo registros de una misma sección o tema (SALUD, EDUCACIÓN, POBLACIÓN, GASTOS, etc.) para facilitar la gestión de los mismos por parte del usuario (BASE.WRT).

Esta base finalmente es convertida al formato .SAV para su procesamiento en el paquete SPSS ó al formato .DBF

BASE DE DATOS DE LA EI2000/1

Archivo.SAV	Secciones que contiene	Página del Cuestionario
R01_00	SECCIÓN 2, PARTE A - VIVIENDA	04 y 05
R02_00	SECCIÓN 1 - CARACTERÍSTICAS DE LA POBLACIÓN	02 y 03
	SECCIÓN 6, PARTE A - RESUMEN DE LAS ACTIVIDADES ECONÓMICAS DE LOS MIEMBROS DEL HOGAR	22 a 24
	SECCIÓN 6, PARTE B - TRABAJO PRINCIPAL AGROPECUARIO Y NO AGROPECUARIO DURANTE LOS ÚLTIMOS 7 DÍAS)	25 a 29
	SECCIÓN 6, PARTE C - TRABAJO SECUNDARIO DURANTE LOS ÚLTIMOS 7 DÍAS	30 a 31
	SECCIÓN 6, PARTE D - SUBEMPLEO Y BÚSQUEDA DE TRABAJO SUPLEMENTARIO	32 y 33
	SECCIÓN 6, PARTE E - INGRESOS DEL TRABAJO Y OTRAS FUENTES	34
	SECCIÓN 6, PARTE F - INGRESOS EXTRAORDINARIOS EN LOS ÚLTIMOS 12 MESES	35
R03_00	SECCIÓN 9, PARTE E - GASTOS POR TRANSFERENCIAS A TERCEROS	62
R04_00	SECCIÓN 2, PARTE B - INVENTARIO DE BIENES DURADEROS	06
R04_00	SECCIÓN 3 - EDUCACIÓN	07 a 11
R05_00	SECCIÓN 4, PARTE A - SALUD (PARA TODOS LOS MIEMBROS)	12 a 15
	SECCIÓN 4, PARTE B - SALUD (PARA MUJERES DE 15 A 49 AÑOS)	16 y 17
	SECCIÓN 5, PARTE A - SALUD (PARA MENORES DE 5 AÑOS)	18 y 20
	SECCIÓN 5, PARTE B - ANTROPOMETRÍA (PARA MENORES DE 5 AÑOS)	21
R06_00	SECCIÓN 7 - ACTIVIDAD INDEPENDIENTE DEL HOGAR	36 a 38
R07_00	SECCIÓN 8, PARTE A - INFORMACIÓN SOBRE LOTES	39
R08_00	SECCIÓN 8, PARTE B - MAQUINARIAS E IMPLEMENTOS AGRÍCOLAS	40
R09_00	SECCIÓN 8, PARTE B - INFORMACIÓN SOBRE VENTA DE MAQUINARIAS E IMPLEMENTOS AGRÍCOLAS	40
R10_00	SECCIÓN 8, PARTE C - ANIMALES	41
R11_00	SECCIÓN 8, PARTE C - PRODUCCIÓN PECUARIA	41
R12_00	SECCIÓN 8, PARTE D - CULTIVOS AGRÍCOLAS	42
R13_00	SECCIÓN 8, PARTE E - PROCESAMIENTO DE PRODUCTOS AGROPECUARIOS Y FORESTALES	43
R14_00	SECCIÓN 8, PARTE F - INSUMOS	43
R15_00	SECCIÓN 8, PARTE G - ALQUILER DE RECURSOS	44
R16_00	SECCIÓN 8, PARTE H - SERVICIOS	44
R17_00	SECCIÓN 9, PARTE A - GASTOS ALIMENTARIOS DEL HOGAR	45 a 58
R18_00	SECCIÓN 9, PARTE A - INFORMACIÓN SOBRE SAL FINA Y GRUESA (CONSUMO Y YODACIÓN)	55
R19_00	SECCIÓN 9, PARTE B - GASTOS ALIMENTARIOS FUERA DEL HOGAR	59
R20_00	SECCIÓN 9, PARTE C - GASTOS NO ALIMENTARIOS DEL HOGAR REFERIDOS A LOS ÚLTIMOS 30 Y 90 DÍAS	60
R21_00	SECCIÓN 9, PARTE D - GASTOS NO ALIMENTARIOS DEL HOGAR REFERIDOS A LOS ÚLTIMOS 12 MESES	61
R22_00	SECCIÓN 10. SEGURIDAD Y VIOLENCIA (EXTRA FAMILIAR)	63
R23_00	SECCIÓN 11. SITUACIONES ADVERSAS	64
R24_00	SECCIÓN 12. OTRAS TRANSACCIONES DEL HOGAR	65
R25A_00	SECCIÓN 13. PARTICIPACIÓN CIUDADANA	66
R25B_00	SECCIÓN 13. PARTICIPACIÓN CIUDADANA	66
R26_00	SECCIÓN 14. EVALUACIÓN DE SERVICIOS	67
R27_00	SECCIÓN 15. EMPLEO INFANTIL	
	PARTE A. TRABAJO REMUNERADO DURANTE LOS ÚLTIMOS 7 DÍAS O ÚLTIMOS 6 MESES	68
	PARTE B. TRABAJO NO REMUNERADO DURANTE LOS ÚLTIMOS 7 DÍAS O ÚLTIMOS 6 MESES	69
	PARTE C. ACTIVIDADES DOMÉSTICAS EN LOS ÚLTIMOS 7 DÍAS	70
R28_00	SECCIÓN 16, EVALUACIÓN DE INDICADORES DE CONSUMO (Preguntas 1 y 2)	71
R29_00	SECCIÓN 16, EVALUACIÓN DE INDICADORES DE CONSUMO (Preguntas 3 a 19)	71 y 72

VIII. Generación de Tabulados

El proceso de generación de tabulados se realiza con el paquete estadístico SPSS, no obstante también se generan tabulados de control en IMPS (Principales resultados) procedentes de las bases en formato ASCII (BASE.ASI)

Los tabulados que se generan para los Principales Resultados son:

- Población por Área de Residencia y Sexo según Actividad Económica.
- Tasa de Actividad, Desempleo y Subocupación, según Área de Residencia y Sexo.
- Población Ocupada por Área de Residencia, según Categoría de Ocupación (Ocupación Principal).
- Población Ocupada por Área de Residencia, según Sectores Económicos (Ocupación Principal).
- Población Ocupada por Área de Residencia, según Nivel de Instrucción.
- Población Económicamente Inactiva por Área de Residencia, según Clasificación.
- Evolución de las Tasas de Desempleo Oculto, Abierto y Total, según Área de Residencia.
- Hogares por Área de Residencia, según cobertura de Servicios.
- Hogares por Sexo del Jefe de Hogar, según Área de Residencia.
- Hogares por Área de Residencia, según Nivel de Instrucción del Jefe/a de Hogar.
- Tasa de Analfabetismo por Sexo, según Área de Residencia.
- Población de 5 años y más de edad por Área de Residencia y Sexo, según Razón de Inasistencia a una Institución de Enseñanza regular.
- Promedio de años de estudio de la Población de 25 años y más de edad por Sexo, según Área de residencia.
- Población de 15 años y más de edad que ha realizado algún curso no formal por Nivel de Instrucción, según Área de Residencia y Sexo.
- Niños y Niñas menores de 5 años por Área de Residencia, según Tipo de Vacunación Recibida.
- Prevalencia de Enfermedades en Niños y Niñas Menores de 5 años por Área de Residencia.
- Población Total por Área de Residencia según Cobertura de Seguro Médico.
- Mujeres en edad fértil por Área de Residencia, según Número de Embarazos.
- Opinión de los Hogares respecto a los Servicios Públicos Recibidos (marzo - mayo 2001)
- Opinión de los Hogares respecto a los Pagos Efectuado por los Servicios Públicos Recibidos (marzo - mayo 2001)
- Opinión de los Hogares que no tienen servicios, respecto a los Servicios que prestan las Empresas Públicas (marzo - mayo 2001)

En la D.G.E.E.C no se han generado desarrollados propios con lenguajes de alto nivel.

IX. Políticas Institucionales de Acceso a las Bases de Datos

Las bases de datos de las Encuestas de Hogares ejecutadas por la DGEEC son accesibles a todos los organismos públicos del país, centros académicos e instituciones del sector privado sin fines de lucro,

nacionales e internacionales. Tienen, por tanto, acceso al uso de las bases de datos todos los individuos que cuenten con respaldo y patrocinio institucional

Los usuarios deben presentar una carta o solicitud dirigida al Director General explicando el tipo de investigación a realizar y los propósitos perseguidos, además de los datos generales de identificación

Los usuarios suscribirán una carta comprometiéndose a lo siguiente:

- No transferir la base de datos a terceros. Otros usuarios pueden solicitarla directamente a la DGEEC.
- Entregar a la DGEEC al menos dos copias de las investigaciones realizadas, tan pronto como estas sean terminadas.
- Asumir la responsabilidad de los resultados y opiniones vertidas.
- Acreditar a la DGEEC como la fuente de datos original.
- Respetar las especificaciones técnicas y el marco conceptual de las Encuestas, en particular las recomendaciones sobre el alcance de la información de las variables para fines de inferencia. La DGEEC no se responsabilizará por las estimaciones realizadas con tamaño de muestra no apropiadas para niveles de inferencia fuera de las recomendadas.

Los usuarios pagarán un cargo por cada base de datos entregada para cubrir los costos de diskettes o CD, copias de la documentación metodológica y gastos administrativos. Para estudiantes e instituciones este cargo tiene el valor equivalente a US\$ 50 y para otros usuarios US\$ 200. Quedan exonerados de pagar este cargo las instituciones y miembros del Comité Interinstitucional del Programa MECOVI y las instituciones copatrocinadoras del Programa.

La DGEEC entregará la base de datos de la Encuesta debidamente consistenciadas y con todos los documentos metodológicos de soporte necesarios para que el usuario pueda operar e interpretar adecuadamente la información (manuales, códigos, diccionario de variables, cuestionario, entre otros).

La base de datos incluirá variables adicionales y estimaciones de indicadores básicos de la encuesta como el ingreso familiar total, el gasto familiar agregado, la PEA, pobreza, entre otros.

La base de datos será entregada en soporte magnético y en un formato de fácil acceso y uso (DBF, SAV, TXT, entre otros).

Los principales resultados son publicados dentro de un plazo máximo de dos meses desde la finalización de los trabajos de campo y los resultados generales dentro del plazo de seis meses.

En la página de la D.G.E.E.C (www.dgeec.gov.py) se pueden encontrar los cuestionarios, diccionarios y tabulados correspondientes a los resultados de las Encuestas de Hogares.

No se han desarrollado aplicaciones que permitan las consultas en línea.

ANEXO 1

Plan de consistencia correspondiente a las secciones de Características de la Población y Vivienda:

Las preguntas con respuestas precodificadas deben ser controladas pregunta por pregunta, que éstas sean los códigos establecidos, es decir, no estén fuera de rango. Ejemplo: Pregunta 3, página 2, Rango de 1 a 6.

ERROR

Para cada pregunta con respuesta abierta, establecer un rango mínimo y máximo probable, acorde a los datos solicitados en la misma. Ejemplo: Pregunta 13, página 27, “guaraníes” mínimo “100000”, máximo “2000000”.

VER

PÁGINA 01**PAG01**

Páginas obligatorias : 01, 02, 04-06, 12, 39. Si período es menor que “19”: también las páginas 45-58, 60, 61, 64, 65.

Si período es mayor que “18” no grabar las páginas 45-61, 63-65.

Si Teléfono de la vivienda (carátula) es distinto a N/T, N/T, S/T, S/T, N/D, N/D, o distinto a los números con o sin paréntesis.

ERROR

PAG01 PAG02 HOMBRES MUJERES TOTAL L02

Controlar que no exista discrepancias entre el número de personas consignado en “carátula”, Pág, 1 según sexo y la cantidad de individuos “hombres” o “mujeres” encontrados en la “lista de personas que habitualmente viven en el hogar”, y que la “suma de hombres y mujeres” coincida con el total de personas miembros.

ERROR

PÁGINA 02**PAG02 P02 P03 P04 P05 P06 P09**

Siempre debe existir un/a jefe/a (P03 = “1”), si existe cónyuge debe ser solamente uno (P03 = “2”).

Que el código de las personas registradas como “cónyuge”, “padre” y “madre” en la columna 5 (P05), sean miembros del hogar (P04 = 1)

Que la persona registrada en pregunta 3 “Relación de parentesco” (P03) como “cónyuge” sea del sexo opuesto al de la persona registrada como “jefe”.

Que el estado civil de ambos sea “casado” o “unido” (P09 = “1” ó “2”).

Que el código de la persona registrada como cónyuge en la pregunta 5 (S05), su sexo (P06) sea opuesto a la persona de la línea en cuestión.

Que el código de la persona registrada como padre en P05, sea de sexo masculino (P06 = “1”) y la registrada como madre de sexo femenino (P06 = “6”).

Que el código de la persona registrada como padre en P05, su edad debe ser “igual o mayor que la edad del hijo + 14” (P02 + 14), y la edad de la madre debe ser “igual o mayor que la edad del hijo + 12” (P02 + 12).

ERROR

PÁGINA 02**PAG02 PAG01 L02 P05 P08 FECHAE**

Controlar que el código de la persona (L02) sea consecutivo.

La edad de la persona se registra desde "0", debiendo controlarse la existencia de "edad ignorada" (999)

Que no exista diferencias entre la registrada como edad cumplida de la persona (P02) y el cálculo realizado con fecha de nacimiento (P08) y fecha de realización de la Encuesta en "carátula".

ERROR

PAG02 PAG16 P02 P05 S16 S18

Si la edad de la persona, página 2, (P02) registrada como "madre" en "Código de identificación" (P05) se encuentra entre 15 y 49 años, ENTONCES,

a) controlar si alguna vez estuvo embarazada en página 16, pregunta 16 (S16 \leq 88).

b) controlar el número de veces que está registrada como "madre" (P05M) la misma persona (mismo código), éste, no debe ser mayor que la cantidad de hijos declarados por esa persona en la pregunta 18 de la página 16 (S18.).

ERROR

PAG02 P02 P05

Si en la pregunta 5 está registrado el código de la madre del individuo, ENTONCES, la diferencia existente entre la edad de la madre y la del hijo no debe ser mayor que "50" (P02 "Edad de la madre" – P02 "Edad del hijo \leq "50"), VER

PÁGINA 03**PAG03 PAG02 P02 P11 P13**

Si edad de la persona es mayor o igual a "1" (P02 \geq "1"), ENTONCES en la página 3 pregunta 11 "Hace 1 año, ¿en qué distrito y departamento vivía?, debe existir datos (P11A = "Distrito", P11B = Departamento", P11C = "Área")

Si edad de la persona es mayor o igual a "5" (P02 \geq "5"), ENTONCES en la página 3 pregunta 13 "Hace 5 años, ¿en qué distrito y departamento vivía?, debe existir datos (P13A = "Distrito", P13B = Departamento", P13C = "Área")

ERROR

PÁGINA 04**PAG04 PAG01 PAG06 TELEF V111 V112**

Si en registro de vivienda tiene "Teléfono celular o línea fija" es igual a "1" (V111 = "1" ó V112 = "1"), ENTONCES, debe tener "Teléfono" en la portada (página 01) y la inversa.

Si en registro de vivienda tiene "Internet" (V113 = 1), ENTONCES, Pág.06, Bienes Duraderos, pregunta 27, "¿Tiene el hogar algunos de los siguientes bienes...?" en la "línea 20, Computadora" debe ser distinto de "cero" (BD27 \neq 0).

ERROR

PÁGINA 04**PAG04 V06 DEPARTAMENTO DISTRITO AREA**

Si departamento es Asunción, código 00, y los siguientes: Dpto. Concepción: distrito Concepción (0101). Dpto. San Pedro: distrito San Estanislao (0208), Departamento Cordillera: distrito Caacupé (0301), distrito Eusebio Ayala (0307), distrito San Bernardino (0316). Departamento Guairá distrito Villarica (0401). Departamento Caaguazú distrito Coronel Oviedo (0501). Departamento Itapúa: distrito Encarnación (0701), distrito Bella Vista (0702), distrito Coronel Bogado (0708). Departamento Misiones: distrito San Juan Bautista (0801). Departamento Paraguari: distrito Paraguari(0901). Departamento Alto Paraná: distrito Ciudad del Este (1001). Departamento Central: distrito Fernando de la Mora (1103), distrito Itá (1105), distrito Lambaré (1107), distrito Limpio (1108), distrito Luque (1109), distrito Mariano Roque Alonso (1110), distrito San Antonio (1113). Distrito San Lorenzo (1114), distrito Villa Elisa (1115). Departamento Ñeembucú :distrito Pilar (1201), distrito Alberdi (1202), Departamento Amambay distrito Pedro Juan Caballero (1301), distrito Bella Vista (1302). Departamento Presidente Hayes: distrito Villa Hayes (1504). Departamento Boquerón: distrito Mariscal Estigarribia (1602) y corresponden al Área Urbana, ENTONCES, “De dónde proviene el agua” debe ser “CORPOSANA “ (V06 = “1”).

VER

PAG04 V06 AREA

Si el área es RURAL, “De donde proviene el agua” debe ser distinto de “CORPOSANA” (V06 <> “1”).
ERROR

Inventario de bienes duraderos

PÁGINA 06**PAG06 LINEA BD28**

Cuando la LINEA del “BIEN” en la pregunta 27 “¿Cuál fue el valor de compra de éste?” es distinto de “98” (BD28 <> “98”), ENTONCES, el valor de compra se compara con los valores mínimos y máximos incluidos en una tabla de precios (Gs) y se emite mensaje si el valor consignado “es menor” o “es mayor” a los valores predeterminados

VER

PAG06 LINEA BD30

Cuando la línea del “BIEN” en la pregunta 29 “Año de compra” tiene “ignorado” (BD29 = “ND”) o en la pregunta 28 “¿Cuál fue el valor de compra” se registra “ignorado” (BD28 = “ND”), ENTONCES, la pregunta 30 “¿Cuánto pagaría hoy por éste, como el suyo, en el estado en que se encuentra?” debe tener respuesta, es decir, datos (BD30 <> “blanco”).

ERROR

ANEXO 2

Variables que intervienen en la definición de Actividad Económica (PEAA)

		Si	No
A02	Trabajó durante los últimos 7 días como empleado, cuenta propia o familiar no remunerado	1	6
A03	Tiene algún trabajo u ocupación como obrero, empleado o cuenta propia aunque no lo haya realizado durante los últimos 7 días	1	6
A04	Trabajó durante los últimos 7 días al menos 1 hora por cuenta propia o patrón en una empresa no agropecuaria ó al menos 7 horas en su chacra o cuidado de sus animales	1	6

A05	Por qué no trabajó los últimos 7 días	
	Despido	1
	Buscó trabajo habiendo trabajado antes	2
	Buscó trabajo por 1a. vez	3
	Demasiado joven	4
	Se dedica exclusivamente a las labores de hogar	5
	Espera comenzar un nuevo trabajo dentro de los próximos 30 días	6
	Mal tiempo	7
	Trabajo temporal	8
	No consigue trabajo	9
	Estudiante	10
	Enfermo	11
	Anciano o discapacitado	12
	Rentista	13
	Jubilado o pensionado	14
	Estuvo en otra situación	15

		Si	No
A12	Si se le ofreciera un trabajo en este momento, trabajaría ?	1	6
A14	Ha hecho algo para conseguir trabajo los últimos 7 días ?	1	6

A15	Por qué no buscó trabajo los últimos 7 días ?	
	No quiso trabajar	1
	No cree poder encontrar trabajo	2
	Se cansó de buscar	3
	No sabe donde consultar	4
	Por salud, enfermo	5
	Por su edad	6
	Por mal tiempo	7
	Ha buscado antes y ahora está esperando noticias	8
	Está ocupado en otras actividades	9
	Otra razón	10

METODOLOGÍA PARA LA INTEGRACIÓN DE BASES DE DATOS DE ENCUESTAS DE HOGARES

INSTITUTO NACIONAL DE ESTADÍSTICA E INFORMÁTICA (INEI)

PERÚ

ÍNDICE

	<u>Página</u>
I. Estrategias para la Captura de Datos	215
1. Descentralización del ingreso de datos.....	215
A. Sistema de control	216
B. Sistema de entrada de datos	217
2. Banco de preguntas y respuestas del sistema de procesamiento a través de Internet....	219
II. Procedimientos para Evaluar la Consistencia de la Información.....	220
Consistencia de los datos de las encuesta.....	220
Consistencia básica.....	220
Consistencia final	220
Tiempo de ejecución de la consistencia básica y la consistencia final.....	224
III. Diseños Computacionales de las Bases de Datos	224
III.1. Software empleado para la captura, validación, procesamiento y explotación de las Bases de datos.....	224
III.1.1 Procesamiento primario de datos.....	224
III.1.2 Procesamiento final de datos	225
III.1.3 Explotación de datos.....	225
III.2 Justificación de los sistemas vigentes frente a otras opciones del mercado	225
IV. Métodos para la Imputación de Datos	226
1. Deflactación.....	226
2. Imputación.....	227
V. Agregación de Variables.....	231
VI. Arquitectura de las Bases de Datos.....	241
VI.1 Estructura y naturaleza de los archivos de datos.....	241
VI.2 Identificación de casos y otras características.....	242
VI.3 Coexistencia de variables	242
VI.4 Generación de variables por hogares.....	242
VII. Generación de Tabulados.....	242
VII.1 Uso de sistemas integrados para el procesamiento de encuestas	242

I. Estrategias para la Captura de Datos

Hasta el año 1996 el procesamiento de la ENAHO se caracterizó por ser fundamentalmente centralizado. A partir del año 1997 se inicia una etapa de cambios progresivos en todos los procesos de la encuesta, a continuación abordaremos brevemente los principales cambios en la ETAPA DE INGRESO DE DATOS (captura) los cuales se vienen desarrollando hasta la actualidad:

1. Descentralización del ingreso de datos

Se toma la decisión institucional de DESCENTRALIZAR el procesamiento del ingreso de datos. El objetivo fue que el ingreso de datos esté lo más cerca posible de los responsables y encargados de la recopilación de información (encuestadores), facilitándose con ello la recuperación directa de la misma.

Con la implementación de este sistema de procesamiento en cada sede, las correcciones son ejecutadas basándose en la verificación primaria de los cuestionarios y con el propio personal de campo. Para ello, se utilizó y mejoró los recursos existentes en las Oficinas Departamentales las que se denominaron Sedes de Proceso (25 sedes), a las Sedes se les equipó con computadoras. La descentralización del trabajo contribuyó a un control más directo y oportuno de los datos recolectados.

Para poder llevar a cabo la implementación de esta tarea se elaboró el **Manual del Procesamiento**, este documento contiene en forma detallada las normas, procedimientos manuales y automáticos para la ejecución del procesamiento de datos en forma descentralizada.

Esta integrado por tres manuales, estos tres tipos de documentos son elaborados con la participación de Especialistas en Informática y de Encuestas. Estos son:

➤ **Guía de procedimiento y uso del sistema de control**

Documento que contiene las normas y procedimientos para administrar el material utilizado en campo (Organización e inventario de cédulas, control de documentos

auxiliares diligenciados y que han sido recepcionados después de la operación de campo), generar indicadores que permitan evaluar algunas actividades del personal de campo (tiempo promedio por visita, número de visitas efectuadas al hogar para completar una entrevista, entre otros) y generar el archivo marco de hogares recepcionados que es utilizado para el control de cobertura durante el Procesamiento de Datos.

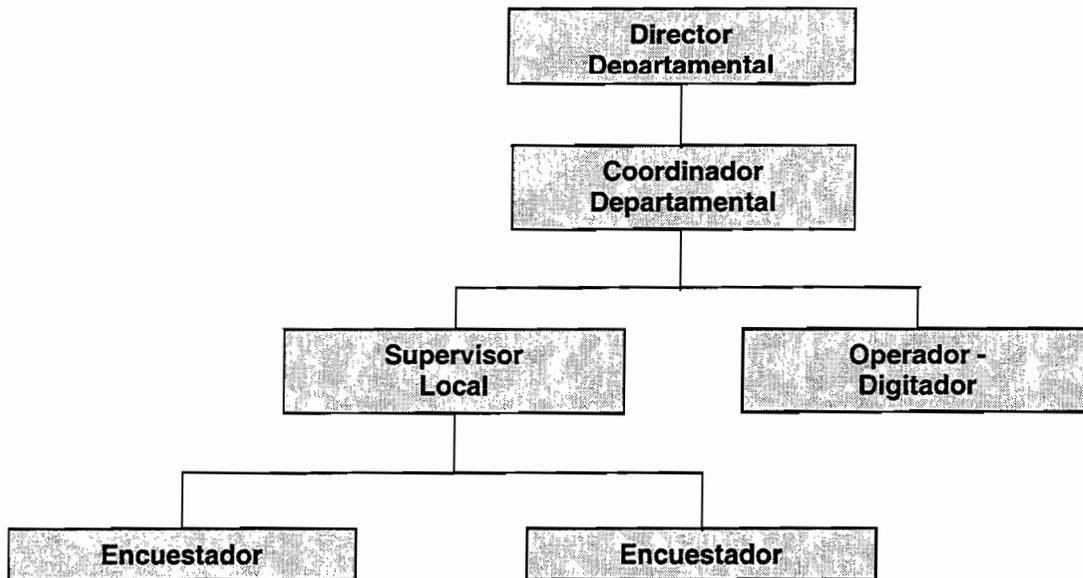
➤ **Sistema de procesamiento básico de datos**

El objetivo de este manual es guiar al Operador Digitador en cada una de las tareas que debe realizar para efectuar la entrada de datos, corrección de la data, control de cobertura, codificación asistida, generación de indicadores de calidad, consistencia, imputación, control de producción y calidad.

➤ **Manual de consultoría**

Documento que contiene las instrucciones que debe seguir el Consultor o la Supervisora Local para resolver las inconsistencias que se producen en la entrada de datos debido a omisiones en campo, relaciones entre variables mal diligenciadas o mal digitadas.

Además se tuvo que implementar en las Oficinas Departamentales la siguiente estructura orgánica funcional para permitir la mejor ejecución de la encuesta:



Director Departamental, es el funcionario de máxima autoridad en la ODEI, responsable de dirigir y administrar las actividades de la Sede de Proceso

Coordinador Departamental, es el funcionario que se encargará de dirigir, supervisar, evaluar y coordinar la ejecución de la ENAHO en el ámbito departamental, asegurando la calidad de la información.

Supervisor Local, es el funcionario debidamente capacitado, responsable y con amplia experiencia en la encuesta, a quien se le confía ejecutar, supervisar y controlar la operación de campo, en un determinado ámbito geográfico.

Encuestador, es la persona que solicita y obtiene información de las viviendas seleccionadas y personas que las ocupan. La calidad de la información dependerá de la eficiencia de su trabajo.

Operador-Digitador, es la persona encargada de ingresar y actualizar la información de las cédulas diligenciadas, en los diferentes sistemas.

Se estableció que dentro de las tareas del Sistema de Procesamiento Básico se debía considerar el control, digitación, cobertura, consistencia, deflactación, imputación y obtención de resultados finales. El sistema de procesamiento básico comprende:

A. Sistema de control

Este sistema realiza el control hasta el nivel de vivienda y hogar de tal manera que es posible identificar de manera automática la vivienda u hogar faltante, reduciendo la actividad manual de identificación de cuestionarios no digitados o extraviados en las otras fases del procesamiento. Además, permite analizar la gestión de la operación de campo, dando elementos para tomar previsiones en los siguientes eventos y sobre todo que permitan mejorar la actividad del encuestador al momento de llenar los cuestionarios.

A través de este sistema se pueden obtener consultas a través de reportes como son:

- Reporte de Inventario de Cuestionarios
- Reporte de Inventario de Cuestionarios y Documentos Auxiliares
- Reporte de Viviendas encuestadas por Resultado Final de la Encuesta
- Reporte de Viviendas y Tiempo Promedio utilizado al concluir las Visitas, según Departamento
- Reporte Total de Hogares encuestados por el número de días en que la encuestadora completa la encuesta.

B. Sistema de entrada de datos

Este sistema permite la Digitación, Modificación, Eliminación y Redigitación de los datos contenidos en los cuestionarios a través de un programa de entrada de “datos inteligente”, es decir, realiza controles de consistencias de rangos y flujos de acuerdo al diseño del cuestionario y las reglas de consistencias preestablecidas. Además se han incorporado controles que permiten medir la calidad de los datos digitados y evaluar los errores cometidos por el encuestador o el digitador, dando elementos para realizar correcciones oportunas.

Para esto último se utiliza la metodología de Control de Calidad por doble digitación y consistencias. La doble digitación consiste en digitar dos veces un dato cuantitativo. Se contabiliza el error cuando el segundo valor es diferente al primer valor digitado. Así mismo el control por consistencia, contabiliza el error cuando no se cumple una regla de consistencia incorporada en el programa de entrada de datos.

Asimismo permite realizar las el control de las coberturas en línea (el digitador no puede omitir personas en ningún módulo, esto se ha logrado haciendo comparación del marco generado en el Sistema de Control versus lo digitado), generar los archivos para realizar la codificación automática, generar los indicadores de avance y calidad, emite listados para consistencia de información, realiza la imputación, genera indicadores de control de calidad y producción de la digitación, permite actualizar las tablas de uso general y proporciona los utilitarios para realizar los backup (copia de respaldo), restaurar archivos y realizar conversiones de archivos a diferentes formatos.

Todo los procesos están integrados por un menú principal. No existen procesos independientes. Los datos del archivo ISSA como en IMPS son similares. No existe ningún proceso adicional en el ambiente IMPS.

En la actualidad, la estructura de datos contiene variables de las preguntas del cuestionario, variables deflactadas, imputadas y variables para datos calculados. Con

este diseño no se modifica por ningún motivo los datos originales provenientes del campo. Es decir en el archivo se tiene datos tal como está en el cuestionario, datos deflactados, datos imputados y datos calculados. Esta última información es para facilitar la explotación del usuario.

A través de la ejecución de la encuesta la data sufre una serie de procesos y acciones específicas que tienen como objetivo el garantizar que los resultados finales sean, coherentes y válidos.

A través del sistema de entrada de datos se pueden generar una serie de reportes de control lo que permite que en cada uno de los departamentos se realice un primer análisis preliminar de la información proveniente de campo (esta tarea se realizó hasta el 2002 IV Trimestre, actualmente en el 2003 se realiza el análisis primario y final de la información en cada departamento).

A continuación se describen algunos de los reportes que se pueden generar a través del sistema de entrada de datos:

□ Cobertura

Proceso automático que permite detectar los cuestionarios, capítulos o preguntas que no han sido digitados y también permite medir la cantidad de registros procesados. Es decir detecta si están completos los datos digitados respecto a los cuestionarios recepcionados y registrados en el Sistema de Control. En este proceso se detectan viviendas omitidas, duplicadas o con identificación errónea. Para tal efecto se debe comparar la información digitada con el marco de conglomerados y número de viviendas seleccionadas.

La emisión de listados, luego de culminada la entrada de datos, permiten identificar por cada conglomerado las viviendas, hogares, capítulos y preguntas que no han sido digitados. Permiten realizar las correcciones ya sea a nivel de digitador o de Supervisor y/o Encuestador. Para recuperar las omisiones restantes se entrega a las encuestadoras y/o supervisores los listados de cobertura a fin de que puedan retornar a los hogares y recuperar la información en forma directa.

Realiza el seguimiento del cumplimiento de la muestra evitando la duplicidad de información. Este proceso, a su vez, implica los siguientes subprocesos:

- **Por tipo de Cuestionario:** identifica omisiones a nivel de cuestionarios y personas. Asimismo da mensajes respecto a miembros que no pertenecen al cuestionario de Productor Agropecuario.
- **Preguntas y Capítulos:** identifica omisiones a nivel de módulos y preguntas. En el proceso de digitación, algunos campos son digitados con valores temporales (alguna cadena de nueves), para permitir la continuación de la digitación. Esta opción tiene – por lo tanto – la responsabilidad de identificar este tipo de situaciones para su verificación ya sea por el digitador como por el encuestador. Algunos de estos valores temporales, son mantenidos y enviados en la data a la Sede Central para continuar con el proceso de consistencia. Esta es una de las situaciones que explican que el proceso de cobertura sea nuevamente ejecutado en la Sede Central.
- **Conglomerados:** esta opción genera tres listados (conglomerados, personas por capítulos, declarados en el marco vs. Digitados).
- **Resumen:** esta opción genera un reporte resumen de cobertura o total de personas. Permite detectar casos de omisión de personas del conglomerado o de la vivienda.

□ Indicadores

Proceso que permite generar una serie de indicadores que permiten medir y evaluar el avance y la calidad de los datos que se van procesando, comparándolos con resultados de encuestas anteriores.

Con esta opción se pueden generar:

- Indicadores de Empleo y Desempleo
- Indicadores de Avance

Este listado nos permite evaluar en un momento determinado (período) la calidad de la información recopilada.

Trimestre	Número de Indicadores
1997 – IV Trimestre	18
1998 – II Trimestre	18
1998 – III Trimestre	23
1998 – IV Trimestre	35
1999 – I Trimestre	27
1999 – II Trimestre	36
1999 – III Trimestre	23
1999 – IV Trimestre	36
2000 – I Trimestre	26
2000 – II Trimestre	36
2001 – IV Trimestre	36
2002 – IV Trimestre	36

Estos indicadores son evaluados por el Coordinador Departamental o Supervisor Local y analizados bajo ciertos parámetros de aceptación que se les ha impartido al personal a través del Manual del Supervisor. La evaluación de estos indicadores se puede efectuar a dos niveles: departamental y por encuestadora.

□ **Consistencia**

El hecho que el procesamiento de la información sea paralelo a la Operación de Campo, permite la generación de un conjunto de Reportes de Control de Calidad, cuya finalidad es detectar posibles errores e inconsistencias de la información recopilada que debe ser corregida en el campo en consulta con el propio informante, lo que ciertamente mejora la calidad de la información registrada.

Esta opción permite generar los reportes siguientes:

- Listado de Preguntas con Alternativas Otros
- Reporte de Balance de Ingresos y Gastos por Hogar
- Reporte de Consumo Percápita
- Reporte de Relación de Precios

2. Banco de preguntas y respuestas del sistema de procesamiento a través de Internet

Se tomo la decisión de aprovechar y hacer uso de las herramientas tecnológicas como es Internet, para el soporte de los sistemas a ser desarrollados:

- Sistema Gestión
- Banco de Preguntas y Respuestas del Sistema Procesamiento
- Banco de Preguntas y Respuestas del Personal de Campo

Esta decisión ha contribuido a propiciar un “acercamiento virtual” a las actividades en ejecución de la encuesta y su correspondiente registro, permitiendo tomar decisiones ante situaciones imprevistas y problemas presentados.

II. Procedimientos para Evaluar la Consistencia de la Información

Objetivo.- Evaluar los procedimientos que se aplican para garantizar la consistencia y confiabilidad de la información recopilada en las encuestas de hogares.

Consistencia de los datos de las encuestas

Comprende las siguientes tareas:

- Consistencia básica
- Consistencia final

Consistencia básica

La descentralización del Procesamiento y la ejecución de esta actividad se realiza en forma paralela a la Operación de Campo, permite la generación de un conjunto de Reportes de Control de Calidad, cuya finalidad es detectar posibles errores e inconsistencias de la información recopilada que debe ser corregida en el campo en consulta con el propio informante, lo que ciertamente mejora la calidad de la información registrada.

➤ Listado de Preguntas con Alternativas Otros

En la actualidad la Supervisora Local revisa este listado y de acuerdo al caso toma acciones correctivas, como codificar la alternativa que se encuentren pre-codificada por que el texto corresponde a alguna de ellas, así como refuerza los conceptos que las Encuestadoras deben tener en cuenta para la utilización de la opción OTROS en las diferentes preguntas.

➤ Reporte de Balance de Ingresos y Gastos por Hogar

Actualmente, este reporte se genera por computadora una vez concluida la digitación de la información del hogar. En este proceso se detectarán los hogares cuyos ingresos y gastos exceden parámetros establecidos por departamentos. La revisión de estos reportes es efectuada por la Encuestadora y Supervisora Local.

➤ Reporte de Consumo Percápita

Actualmente este es revisado por la Supervisora Local luego de concluida la digitación del cuestionario y tiene por objetivo revisar aquellos consumos percápita diario en gramos que superan las dos desviaciones estándar del promedio registrado en la Encuesta de Condiciones de Vida efectuada en 1997.

➤ Reporte de Relación de Precios

Actualmente este reporte es revisado por la Supervisora Local luego de concluida la digitación del cuestionario y tiene por objetivo verificar el precio del producto comprado por kilo o litro que sea diferente a $\pm 15\%$ de acuerdo al promedio mensual de los principales productos que conforman la canasta familiar.

Consistencia final

Hasta el año 2002 la consistencia final se realizaba en forma centralizada en la ciudad de Lima, para ello luego de terminada la digitación (por periodo) en cada una de las ODEI's, los archivos con la información de los cuestionarios digitados se remitía a la Sede Central en donde se daba inicio a esta segunda etapa,

que consistía en centralizar toda la información para realizar la consistencia final mediante el cruce de las variables. Actualmente en cada ODEI se viene realizando este procedimiento.

En esta labor se hace uso del software ISSA-X con su aplicación de menús que se han venido implementando con el propósito de realizar un mejor control de calidad de la información, por ejemplo tenemos que el análisis se ha complementado con el uso del software SPSS, es el mismo sistema de entrada de datos el que realiza la conversión de los archivos a este paquete.

Esta tarea además comprende actividades como son la tabulación y las técnicas especificadas que sirvan para ordenar y analizar la información recopilada en los cuestionarios de la encuesta, mediante el uso de computadoras electrónicas y Software Estadísticos con la finalidad de obtener los resultados de acuerdo al plan de tabulaciones.

La consistencia final comprende las siguientes tareas:

1. Codificación de Preguntas Abiertas

El proceso de codificación en un primer momento se realizó de manera manual, es decir, existían equipos de trabajo especializados para realizar esta labor.

La mejora ha consistido en la elaboración de un programa "diccionario" basado en expresiones literales contenidas en experiencias anteriores, el programa compara si el texto digitado coincide con alguna expresión del "diccionario", si coincide le asigna su código de acuerdo al Anexo de Código de Ocupaciones, Código de Profesiones y Oficios y Código de Actividades Económicas correspondiente, de no coincidir el programa contempla la emisión de un reporte de verificación, es decir, este proceso pasa por 2 etapas:

a. Codificación Automática: Proceso automático, en el cual se asigna los códigos de ocupación y rama de actividad. Finalizado este proceso se realiza la verificación de la codificación automática, proceso el cual genera un reporte en el que se identifica las ocupaciones y ramas que no fueron codificadas por no guardar relación, para su posterior Codificación Asistida.

b. Codificación Asistida: Consiste en el análisis y corrección en la base de datos de los Códigos de Ocupación y Rama de Actividad que no fueron considerados en el proceso de la Codificación Automática.

Actualmente se ha incrementado el porcentaje de codificación automática de las variables de Rama de Actividad y Ocupaciones, se tiene el 60% aproximadamente y se espera llegar a un 70% de casos codificados.

Entre las preguntas ha ser codificadas tenemos las siguientes:

P505: ¿Cuál es la ocupación principal que desempeñó?

P506: ¿A qué se dedica el negocio, organismo o empresa en la que trabajó?

P580: ¿Cuál es (fue) el nombre del curso o programas de capacitación?

P585: ¿Cuál es ese oficio?

2. Análisis de la Variable OTROS

Consiste en la revisión de las preguntas donde existen las alternativas "Otros" para verificar que esta no corresponda a una alternativa pre-codificada. También se verificará que este bien redactado y el porcentaje por pregunta para ver si es necesario recodificar la alternativa (se revisa todas las preguntas cuya alternativa Otro supere el 5 % de su valor).

3. Revisión de Marginales

Consiste en la revisión de valores extremos, estructuras porcentuales y flujos de las todas las preguntas de los cuestionarios en estudio.

4. Revisión de valores extremos de las Variables Numéricas (Análisis Exploratorio)

Consiste en la revisión de los ingresos o gastos y valores altos o bajos de las preguntas numéricas. Esta revisión se realiza a través del Software Estadístico SPSS.

5. Revisión de cuadros básicos de análisis

En esta etapa se verifica si existen inconsistencias entre preguntas mediante el cruce de variables. Esta revisión se realiza a través del Software Estadístico SPSS.

Ejemplo:

- **Cuestionario Enaho.01**
 - Capítulo 100: Características de la Vivienda y del Hogar
 - Tipo de Vivienda por Tenencia de la vivienda
 - Tipo de Vivienda por Abastecimiento de agua de su hogar procede de:
 - Tipo de Vivienda por Abastecimiento de agua de su hogar procede de, según último gasto mensual por consumo.
 - Tipo de Vivienda por El tipo de Combustible que usan en el hogar para cocinar sus alimentos, según último gasto mensual por consumo.
 - Su hogar tiene, por último gasto mensual por consumo.
 - Capítulo 200: Características de los Miembros del Hogar
 - Estado Civil por edad
 - Edad por Relación de Parentesco
 - Capítulo 600: Gastos en Alimentos y Bebidas
 - Productos por Unidad de Medida, Comprados y Obtenidos (por producto)
 - Productos por Frecuencia y Gasto de la compra (por producto)
 - Capítulo 700: Programas Sociales
 - Conoce Ud. Algunas formas de ayuda del gobierno o instituciones privadas tales como por Organismos o Institución ejecuta dicho programa social
 - Capítulo 800: Otras Transacciones
 - Qué institución o persona le(s) otorgó el préstamo por cuál es el último pago mensual que realizó por la devolución del préstamo recibido.
 - Qué institución o persona le(s) otorgó el préstamo por cuál es el último pago mensual que realizó por cuál es la tasa de interés mensual que paga por el préstamo recibido.
 - Cuál es el último pago mensual que realizó por la devolución del préstamo recibido por Cuánto les queda por pagar por el préstamo recibido.
- **Cuestionario Enaho.01A**
 - Capítulo 300: Educación

- Edad por Nivel educativo aprobado
- Edad por Nivel educativo asistido
- Nivel educativo aprobado por Sabe leer y escribir
- Nivel educativo asistido por En el presente año recibió ayuda del Estado
- Centro de Estudios al que asiste por En el presente año recibió ayuda del Estado
- Centro de Estudios al que asiste por En los últimos 12 meses, ¿A cuánto asciende aproximadamente el gasto por:
- Centro de Estudios al que asiste por ¿A cuánto asciende aproximadamente el último gasto mensual en estudios por:
- Edad por ¿Cuál es la principal razón por la que no asiste a algún centro de enseñanza regular?
- o Capítulo 400: Salud
- Edad por ¿En qué día, mes y año nació?
- Sexo por ¿Cuál fue la enfermedad o accidente que tuvo?
- Edad por ¿Cuál fue la enfermedad o accidente que tuvo?
- ¿Cuál fue la enfermedad o accidente que tuvo? Por ¿Dónde se llevó a cabo la consulta o atención de salud?
- ¿Dónde se llevó a cabo la consulta o atención de salud? Por ¿Quién le atendió durante la consulta?
- ¿Dónde se llevó a cabo la consulta o atención de salud? Por Recibió los servicios de: (Consulta, Análisis, Rayos X, Otro examen)
- ¿Dónde se llevó a cabo la consulta o atención de salud? Por ¿Cuánto gastó en medicinas?
- ¿Dónde estuvo hospitalizado? Por ¿Cuántos días estuvo hospitalizado y cuál fue el monto total gastado?
- o Capítulo 500: Empleo e Ingreso
- La semana pasada, del ... al ... ¿Tuvo Ud. Algún trabajo? Por Categoría de Ocupación
- Aunque no trabajó la semana pasada, ¿Tiene algún empleo fijo al que próximamente volverá? Por Categoría de Ocupación
- Aunque no trabajó la semana pasada, ¿Tiene algún negocio propio al que próximamente volverá? Por Categoría de Ocupación
- La semana pasada, realizó alguna actividad al menos una hora para obtener ingresos o en especie, como: Por Categoría de Ocupación
- ¿Cuál es la ocupación principal que desempeñó? Por ¿A qué se dedica el negocio, organismo o empresa en la que trabajó?
- ¿Cuál es la ocupación principal que desempeñó? Por Edad
- ¿Cuál es la ocupación principal que desempeñó? Por Sexo
- ¿Cuál es la ocupación principal que desempeñó? Por Nivel educativo aprobado
- ¿Cuál es la ocupación principal que desempeñó? Por Categoría de Ocupación

- ¿A qué se dedica el negocio, organismo o empresa en la que trabajó? Por En su ocupación principal Ud. Trabajó para:
 - Categoría de ocupación por En su ocupación principal Ud. Trabajó para:
 - Categoría de ocupación por ¿El tipo de pago o ingreso que Ud. Recibe en su ocupación principal es:
 - Categoría de Ocupación (Actividad Principal) por Categoría de Ocupación (Actividad Secundaria).
 - Categoría de Ocupación por ¿Cuánto fue su ingreso total en el(la) ...anterior, incluyendo horas extras, bonificaciones, pago por concepto de refrigerio, movilidad, comisiones, etc.?
 - Categoría de Ocupación por ¿Cuál es el promedio del ingreso líquido que Ud. Recibió en los dos meses anteriores,?
 - Categoría de Ocupación por En su ocupación principal, ¿Cuál fue su ganancia neta en el mes anterior?
 - Categoría de Ocupación por ¿En su ocupación principal con que frecuencia obtiene sus ingresos?

Tiempo de ejecución de la consistencia básica y la consistencia final

Desde el año 1995 hasta el año 2000, el tiempo de ejecución de estas tareas superaba los 6 meses. En los años 2001 y 2002, el tiempo necesario se redujo a 3 meses después de concluida la operación de campo. Actualmente se viene ejecutando la ENAHO continua desde mayo del presente año y el tiempo necesario para efectuar estas tareas es de 15 días después de concluida la operación de campo del mes correspondiente.

III. Diseños Computacionales de las Bases de Datos

Cada una de las tareas de procesamiento presenta cierto nivel de complejidad, que van a influir al final en la elección de la herramienta software a ser empleada, En este punto, mencionaremos brevemente la tareas realizadas y luego el software empleado.

III.1. Software empleado para la Captura, Validación, Procesamiento y Explotación de las Bases de datos

III.1.1 Procesamiento Primario de Datos

El procesamiento primario de datos comprende las siguientes tareas:

- Digitación de Carátulas.
- Digitación de Cuestionarios.
- Cobertura o Completitud de Datos.
- Consistencia Primaria o Calidad en los Datos Digitados.

Para la ejecución de estas tareas se tienen implementados los siguientes sistemas:

- Sistema de Control, en donde se efectúa el proceso de digitación de carátulas del Cuestionario Principal Enaho.01.
- Sistema de Procesamiento Básico, en donde se efectúa los procesos de digitación del contenido de los Cuestionarios Enaho, Cobertura, Consistencia Primaria.

En el Sistema de Control, se registran las carátulas del cuestionario principal del hogar, con la finalidad de: generar un marco de digitación, que indique los casos validos para ser digitados posteriormente.

El Sistema de Control es una aplicación desarrollada íntegramente en Visual Basic 6.0 que genera datos en una base de datos ACCESS 7.0.

En el Sistema de Procesamiento Básico, se digitan el contenido de todos los cuestionarios de la encuesta (ENAHO 01, 01A, 01B, 02, 03, 04), que han sido

registrado en el Sistema de Control. Para ello emplea el marco de digitación generado por el Sistema de Control. Es decir, sólo se podrá digitar el contenido de los cuestionarios cuya carátula se haya registrado previamente en el Sistema de Control.

Una vez digitados los cuestionarios correspondientes a un período de procesamiento, la siguiente tarea es verificar la completitud de los datos o cobertura, en la cual se determina que los datos de todos los hogares y de todas las personas estén digitadas, además se verifica si proceden las omisiones de datos y las omisiones de módulos.

Cuando los datos están digitados y verificada su completitud, se procede a verificar la Calidad de los mismos o Consistencia Primaria (Relaciones de Precios, Balances de Gastos e Ingresos del Hogar, Consumo Percápita)

El Sistema de Procesamiento Básico consta de un menú de opciones desarrollado en FoxPro 2.6 que tiene como finalidad invocar cada una de las tareas implementadas mediante aplicaciones desarrolladas en ISSA-X.

III.1.2 Procesamiento Final de Datos

El procesamiento final se inicia con la secuencia iterativa de los siguientes procesos:

- Conversión de los archivos de datos ISSA (en formato texto) hacia archivos en SPSS (formato SAV) .
- Detección de inconsistencias, para lo cual se emplean archivos de sintaxis en SPSS (secuencias de instrucciones) que al ser ejecutados generan cuadros de resultados, en donde se identifican los casos que presentan inconsistencias
- Correcciones de los datos en los Archivos ISSA

Una vez que los datos están libres de inconsistencias, se ejecuta la segunda parte del procesamiento final, la cual esta íntegramente implementada en ISSA-X y que consiste en:

- Trimestralización y Deflactación de Variables (Ingresos y Gastos).
- Imputación de Variables (Horas, Gastos, e Ingresos), empleando matrices de valores promedios.
- Generación de Variables Calculadas a Nivel de Hogar.

III.1.3 Explotación de Datos

Finalmente, los archivos de datos finales son convertidos de formato ISSA-X a formato SPSS para su manejo por parte del usuario. Estos archivos, como se verá más adelante, guardan semejanza en su estructura con el diseño de cada módulo de los cuestionarios de la encuesta.

III.2 Justificación de los Sistemas Vigentes frente a otras opciones del Mercado

Siendo la Encuesta Nacional de Hogares una encuesta laboriosa en procesamiento, debido más que nada a la cantidad de preguntas a procesar por cuestionario. Y a lo limitante que puede ser el parque actual de

computadoras, han justificado en parte que el ISSA-X sea empleado como entorno de desarrollo y procesamiento.

Entornos de desarrollo de alto nivel como Visual FoxPro o Visual Basic que si bien son entornos de desarrollo más amigables, presentan el problema de limitación en el manejo de pantallas.

Una alternativa un tanto a considerar estaría dada por un entorno de procesamiento primario descentralizado desarrollado MS ACCESS, aunque aquí habría que estandarizar la versión a utilizar de las diversas que existen en el mercado (ACCESS 97, ACCES 2000 , ACCESS XP).

IV. Métodos para la Imputación de Datos

¿Cómo se trata la no respuesta total y parcial?

Las preguntas que no cuentan con respuesta (Preguntas omitidas) durante el diligenciamiento de los cuestionarios, en el momento de la entrada de datos se le asigna un determinado valor.

Sólo en el caso de los alimentos consumidos dentro del hogar que no ocasionan gasto, como es el caso del autoconsumo, autosuministro, como parte de pago, donaciones, regalos y otros; se imputan valores de gasto en consumo a partir de los precios implícitos que provienen del gasto efectuado en alimentos comprados por los hogares.

¿Cuáles criterios se aplican para la imputación de datos faltantes?

Para realizar la imputación de los datos faltantes se emplean 2 etapas:

1. Deflactación

Actualmente se realiza primero la Deflactación y trimestralización de los valores monetarios y luego las imputaciones. La deflactación y trimestralización, se obtiene tomando como base la fecha de referencia y mediante la aplicación de un índice de precios se elimina el efecto de los precios en el período de análisis. Ejemplo:

Sea

t_d = día de inicio de la semana de referencia

t_m = mes de inicio de la semana de referencia

FECHA DE REFERENCIA	+ DEFLACTOR (IPC)
Fecha del Capítulo 500	
(Diario = 1) P529 (1:6) > 0 * 25 * 3	t_m
(Semanal = 2) P529 (1:6) > 0 * 4.33 * 3	t_m
(Quincenal = 3), si $t_d = 1 - 15$ P529 (1:6) > 0 * 6	t_{m-1}
(Quincenal = 3), si $t_d = 16 - 31$ P529 (1:6) > 0 * 6	t_m
(Mensual = 4) P529 (1:6) > 0 * 3	t_{m-1}
	$(t_{m-1} + t_{m-2})$

FECHA DE REFERENCIA	+ DEFLACTOR (IPC)
(Bimestral = 5) P529 (1:6) > 0 * 3/2	2
(Trimestral = 6) P529 (1:6) > 0	$\frac{(t_{m-1} + t_{m-2} + t_{m-3})}{3}$
(Semestral = 7) P529 (1:6) > 0 + 2	$\frac{(t_{m-1} + t_{m-2} + \dots + t_{m-6})}{6}$
(Anual - Rara vez = 8) P529 (1:6) > 0 + 4	$\frac{(t_{m-1} + t_{m-2} + \dots + t_{m-12})}{12}$

2. Imputación

El procedimiento empleado es conocido con el nombre de “Imputación por Medias”, es decir, se construyen matrices de valores promedios (Trimestrales) para la asignación de datos faltantes y posteriormente se emiten los listados que contengan las identificaciones del registro de hogares imputados, así como indicadores sobre el porcentaje imputado a nivel de cada pregunta.

En la Encuesta Nacional de Hogares 2002 - IV Trimestre se imputaron las preguntas siguientes:

Pregunta 105 y 106: Alquileres

Se construyó la matriz de valores promedios en base a las variables:

- Tipo de Vivienda: Pregunta 101
- El Material Predominante en las Paredes Exteriores es: Pregunta 102
- Sin contar Baño, Cocina, Pasadizos ni Garaje ¿Cuántas Habitaciones en total tiene la Vivienda?: Pregunta 104
- La Vivienda que ocupa su lugar es, Alquilada: Pregunta 105-A y 105-B
- ¿Cuál es el último grado o año de estudios que aprobó? El Jefe del Hogar: Pregunta 301-A
- Area (urbana, rural)
- Dominio de estudio

Pregunta 117: Último gasto por consumo de agua

Se construyó la matriz de valores promedios en base a las variables:

- Abastecimiento de agua en su hogar procede de: Pregunta 110
- Número de miembros por Hogar: Pregunta 204=1
- Estratos Socioeconómicos para Lima Metropolitana, por Conglomerados
- Area (urbana, rural)
- Dominio de estudio

Pregunta 117: Último gasto por consumo de electricidad

Se construyó la matriz de valores promedios en base a las variables:

- Sin contar Baño, Cocina, Pasadizos ni Garaje ¿Cuántas Habitaciones en total tiene la Vivienda?: Pregunta 104

- Número de miembros por Hogar: Pregunta 204=1
- Número de Electrodomésticos: Pregunta 612(i)=1 y 612-A
- Dominio de estudio

Pregunta 117: Último gasto por consumo de combustible:

Se construyó la matriz de valores promedios en base a las variables:

- Area (Urbana, rural)
- Pregunta ¿Cuál es el tipo de alumbrado y combustible que utiliza su hogar?: Pregunta 112 y 113
- Promedio del gasto per cápita por alumbrado o combustible
- Dominio de estudio

Pregunta 117: Último gasto por consumo en comunicaciones:

Se construyó la matriz de valores promedios en base a las variables:

- Su hogar tiene: Teléfono, Celular, Beeper (Pregunta 114)
- Número de miembros por Hogar: Pregunta 204=1
- ¿Cuál es el último grado o año de estudios que aprobó? El Jefe del Hogar: Pregunta 301-A
- Dominio de estudio

Pregunta 601: Gasto en alimentos y bebidas consumidos dentro del hogar

Se homogenizaron a kilos o litros las unidades de medida en que fueron obtenidos los productos adquiridos por el hogar.

Se estandarizaron las frecuencias en que se declararon las cantidades y gastos de los productos comprados y obtenidos llevándolas a valores trimestrales. Se obtuvo la matriz de la mediana de precios de acuerdo las variables:

- Producto: Pregunta 601
- Departamento
- Area (Urbana, rural)

Pregunta 602: Gasto en alimentos obtenidos de Instituciones Benéficas para consumir dentro del hogar

Generar la matriz para la imputación de gastos alimentos para consumir dentro del hogar obtenidos de instituciones benéficas en base al valor promedio de acuerdo al cruce de las variables:

- En los últimos 15 días, del ... al ... Ud. o algún miembro de este hogar obtuvieron, compraron o le regalaron alimentos para consumir en el hogar?: Pregunta 602
- ¿Cuántas veces por semana? Pregunta 602-A (i) * 4.33 * 3
- Generalmente ¿Cuántas raciones consumieron por vez?:
- Pregunta 602-B * 602-A * 4.33 * 3
- Generalmente ¿Cuánto pagó cada vez?: Pregunta 602-E
- Dominio de estudio

Pregunta 603 a 611: Otros gasto del hogar

Generar la matriz para la imputación de gastos percápita de acuerdo al cruce de las variables:

- Rubro de gasto: En el mes anterior, del ... al ... Ud. o algún miembro de este hogar obtuvieron, consiguieron, compraron o le regalaron lo siguiente: Preguntas 603-611
- ¿Cuánto fue el monto total por la compra?
- ¿En cuánto estima el valor del producto o servicio?
- Numero de miembros del hogar: Pregunta 204=1
- Dominio de estudio

Pregunta 612: Gasto en equipamiento del hogar

Generar la matriz para la imputación de gastos: Equipamiento del hogar en base al promedio de acuerdo al cruce de las variables:

- Su hogar tiene: Pregunta 612(i) = 1
- ¿Cuánto paga?: Pregunta 612-G
- ¿Con qué frecuencia paga?: Pregunta 612-F
- Dominio de estudio

Pregunta 311 y 312: Gastos en educación

Generar la matriz para la imputación de gastos en educación en base al valor promedio de acuerdo al cruce de las variables:

- ¿Cuál es el último grado o año de estudios que aprobó? El Jefe del Hogar: Pregunta 301-A
- ¿Cuál es el grado o año de estudios al que asiste?: Pregunta 308-A
- Centro de estudios (Estatad, Particular): Pregunta 308-D
- Dominio de estudio

Pregunta 416 y 418: Gastos en servicios de salud

Generar la matriz para la imputación de gastos por salud de acuerdo al cruce de las variables:

- Rubro de gasto: Ud. Recibió: Pregunta 414
- Edad : Pregunta 208 tenemos Menores de 1 año, De 1 a 4 años, De 5 a 14 años, De 15 a 50 años, De 51 a 65 años y De 65 y más.
- ¿Cuánto fue el monto total por la compra o servicio?
- ¿En cuánto estima el valor de?
- Dominio de estudio

Pregunta 513 y 518: Horas

Generar la matriz de imputación; en base al valor promedio de acuerdo al cruce de las variables de los perceptores:

- Sexo: Pregunta 207
- Edad: Pregunta 208 según grupos siguientes: De 14 a 24 años, De 25 a 44 años, De 45 a 64 años y De 65 y más

- Rama de actividad: Pregunta 506
- Categoría de Ocupación: Pregunta 507
- Dominio de estudio

Pregunta 524, 530, 538, 541, 544: Ingresos

Generar la matriz en base al factor promedio monetario por hora diaria a tres decimales del cruce de las variables:

- Sexo: Pregunta 207
- Edad: Pregunta 208, De 14 a 24 años, De 25 a 44 años, De 45 a 64 años, De 65 y más.
- Rama de actividad: Pregunta 506
- Categoría de Ocupación: Pregunta 507
- Dominio de estudio

Pregunta 559: Gastos en alimentos consumidos fuera del hogar

Se estandarizaron las frecuencias de los y gastos efectuados al trimestre.

Se obtiene la matriz de gasto promedio por vez y persona en base a:

- La semana pasada el alimento o bebida que consumió fuera de su hogar correspondió a: Desayuno, Almuerzo, Cena y Otro

Pregunta 560: Otros Gastos consumidos fuera del hogar (transporte público y teléfono público)

Se estandarizaron las frecuencias de los y gastos efectuados al trimestre.

Se obtiene la matriz de gasto promedio por persona en base a:

- La semana pasada hizo Ud. Uso del servicio correspondiente a: Transporte público urbano e interurbano, Teléfono público (Usando tarjeta o moneda).

¿Cómo se evalúan los resultados con datos imputados?

Los resultados de las variables imputadas se evalúan de la manera siguiente :

- Primero, que las variables imputadas no tengan un porcentaje alto con respecto a las variables que cuentan con información.
- Segundo, que las variables imputadas no distorsionen los resultados que se proporcionan con las variables originales.
- Tercero, se comparan las variables imputadas con variables de Encuestas anteriores. Ejemplo:

ACTIVIDAD PRINCIPAL		2001-IV	2002-IV
TRABAJO DEPENDIENTE			
P524A	Ingreso bruto dependiente	0.81	0.25
P524B	Desc de ley	0.90	0.26
P524C	Otros Descuentos	0.72	0.22
P524D	Ingreso neto dependiente	0.57	0.15
TRABAJO INDEPENDIENTE			
P530	Ingreso independiente	3.50	0.79

¿Cómo se documentan los procesos?

Para que los usuarios conozcan con mayor detalle cuales son los procedimientos de la construcción de las matrices para las imputaciones de Horas, Ingresos y Gastos de los cuestionarios de la Encuesta de Hogares, se encuentran debidamente documentados en Archivos de formato Word y cuando se entregan a los usuarios se les remite en Archivos de formato *.pdf

¿Se informa a los usuarios de las variables que han sido intervenidas?

Si, se informa. Los usuarios para que tengan mayor facilidad en el manejo de las Bases de Datos se les proporciona los Diccionarios de Variables, en estos documentos van definidas las variables originales y las variables agregadas (variables deflactadas e imputadas).

¿Se conservan las bases de datos que han sido modificadas?

Las bases de datos no son modificadas; es decir, para evitar que los datos obtenidos en la etapa de la Operación de Campo, no sean modificados en el momento de los procesos de la Deflactación e Imputación de variables se asignan nuevas variables. Ejemplo:

Pgta. 524: ¿Cuánto fue el ingreso total en el (la) , anterior, incluyendo horas extras, bonificaciones, pago por concepto de refrigerio, movilidad, comisiones, etc.?

VARIABLES ORIGINALES	VARIABLES DEFLACTADAS	VARIABLES IMPUTADAS
P524a1: Ingreso Total P524b1: Dsctos. de Ley P524c1: Otros Dsctos. P524d1: Ingreso Líquido	D524a1: Ingreso Total D524b1: Dsctos. de Ley D524c1: Otros Dsctos. D524d1: Ingreso Líquido	I524a1: Ingreso Total I524b1: Dsctos. de Ley I524c1: Otros Dsctos. I524d1: Ingreso Líquido

V. Agregación de Variables

La Encuesta Nacional de Hogares realiza todo un proceso de agregación y/o integración de variables para el cálculo de las variables de ingreso, gastos, consumo, condición de ocupación, entre otras.

Todas las variables agregadas y/o calculadas sí se integran en la base de datos; en el proceso de unión de registros se integran las siguientes variables:

Condición de ocupación (OCU500):

En el Módulo de Empleo e Ingresos, las bases de datos tienen integradas las variables de: Factor de expansión, Ingreso de la persona, condición de actividad, puesto que a través de estas variables se tiene un panorama amplio para analizar los diferentes aspectos que se pueden tratar en una encuesta de este tipo. Cuando un usuario adquiere cualquiera de estas bases, estas siempre van acompañadas de su correspondiente diccionario de datos, en el cual se describen los códigos y etiquetas o significados de los mismos

1 : Ocupados

2 : Desocupados Abiertos

3 : Desempleo Oculto

4 : Inactivos

0 : Omisión del Capítulo

Ingresos y gastos agregados**Ingreso no agrícola:**

1. En su ocupación principal ¿Cual fue el ingreso total que percibió de:
 - Sueldo bruto en el mes anterior?
 - Salario bruto la semana pasada?
 - Otros ingresos no permanentes?
2. Además del ingreso anterior ¿recibió alimentos, vestidos, transporte, vivienda, etc. como parte de pago por su trabajo?
 - Si No
3. En cuanto estimaría Ud. el pago por:
 - ¿Alimentos?
 - ¿Vestido y calzado?
 - ¿Transporte?
 - ¿Vivienda?
 - ¿Salud?
 - ¿Otro?
4. En su ocupación principal ¿Cual fue su ganancia neta en el mes anterior?
5. De los bienes producidos por el hogar y/o de los productos adquiridos con fines comerciales ¿Utilizan productos para su consumo?
 - Si
 - No
6. ¿En cuanto estima Ud. el valor de los productos utilizados para su consumo en el mes anterior?
7. En sus ocupaciones secundarias ¿Cuál fue su ingreso en el mes anterior?
8. En el mes derecibió Ud. ingresos por concepto de:
 - ¿Pensión de jubilación/cesantía?
 - ¿Remesas periódicas de otros hogares?
 - ¿Pensión de divorcio y separación?
 - ¿Pensión por alimentación?
 - ¿Pensión por viudez, orfandad y sobrevivencia?
 - ¿Otro?
9. ¿Cuánto recibió en total por dichos conceptos?
10. En el mes derecibió Ud. ingresos por concepto de:
 - ¿Utilidades?
 - ¿Intereses por depósitos en bancos, cooperativas?
 - ¿Dividendos de acciones, bonos, etc.?
 - ¿Arrendamiento de casas y maquinaria?
 - ¿Arrendamiento de tierras agrícolas?
 - ¿Otro?

11. ¿Cuánto recibió en total por dichos conceptos?
12. En los últimos tres meses, dearecibió Ud. ingresos por concepto de:
 - ¿Indemnizaciones?
 - ¿Seguro de accidentes o vejez?
 - ¿Herencia?
 - ¿Escolaridad?
 - ¿Gratificaciones?
 - ¿Juegos de azar?
 - ¿Otros ocasionales?
13. ¿Cuánto recibió en total por dichos conceptos?

Ingreso agrícola

El ingreso proveniente de actividades agrícolas es investigado con un cuestionario aparte, dirigido a cada productor agropecuario Enaho 02: "Ingreso del Productor Agropecuario".

Se formulaban varias preguntas correspondientes a los capítulos: producción agrícola, subproductos agrícolas, producción forestal, gastos en actividades agrícolas y/o forestales, producción pecuaria, subproductos pecuarios; y, gastos en actividades pecuarias. Se solicitaba información sobre el destino de la producción tal como: venta, consumo del hogar, trueque, sub producción y otros.

En cuanto al gasto en actividades agrícolas y/o forestales, se investiga el gasto efectuado en:

- ☉ Semillas.
- ☉ Abonos y fertilizantes.
- ☉ Pesticidas, insecticidas fungicidas, etc.
- ☉ Compra de sacos, canastas, cajones u otros envases para sus productos.
- ☉ Transporte.
- ☉ Almacenamiento de productos.
- ☉ Pago a jornaleros o peones.
- ☉ Arrendamiento de tierras.
- ☉ Elaboración de subproductos.

En cuanto al gastos en actividades pecuarios se desagregan en varios ítems, se investiga lo siguiente:

- ☉ Gastos en alimentos para animales.
- ☉ Compra de animales.
- ☉ Servicios veterinarios.
- ☉ Productos veterinarios.
- ☉ Pago a jornaleros o peones.
- ☉ Elaboración de subproductos.
- ☉ Otros gastos.

Gastos del hogar

El informante sobre los gastos del hogar es el Jefe o el Ama de Casa. Se investigan los siguientes rubros de gastos:

- ② Alimentos, bebidas y tabaco adquiridos, se desagregaron en 27 ítems y el período de referencia correspondió a los últimos 15 días.
- ② Alimentos, bebidas y tabaco autoconsumidos y autosuministrados, se desagregaron en 25 ítems y el período de referencia correspondió a los últimos 15 días.
- ② Los gastos en mantenimiento de la vivienda, se desagregaron en 5 ítems, con un periodo de referencia en los últimos 15 días.
- ② Los gastos en transportes y comunicaciones, se desagregaron en 6 ítems, con un periodo de referencia en los últimos 15 días.
- ② Los gastos en salud y servicios médicos, se desagregaron en 5 ítems, con un periodo de referencia del mes anterior.
- ② Los gastos en esparcimiento, diversión, servicios de cultura y enseñanza, se desagregaron en 5 ítems, con un periodo de referencia del mes anterior.
- ② Los gastos en vestido y calzado, se desagregaron en 6 ítems, con un periodo de referencia en los últimos 3 meses.
- ② Los gastos en muebles y enseres, se desagregaron en 5 ítems, con un periodo de referencia en los últimos 3 meses.
- ② Otros bienes y servicios, se desagregaron en 4 ítems, con un periodo de referencia en los últimos 3 meses.

Para cada ítems se solicita el gasto efectuado por el hogar en el período de referencia. En cuanto al autoconsumo o autosuministro se solicita el valor estimado en el consumo de dichos productos. Además se solicita el lugar donde se adquirió el producto, la frecuencia de adquisición, la cantidad adquirida, la unidad de medida y el monto del gasto efectuado.

• **Ingreso monetario**

Está referido al nivel de ingreso monetario de los trabajadores en su empleo o ocupación, teniendo en cuenta que representa la mayor proporción en el ingreso del hogar. Permitirá clasificar a los subempleados invisibles por ingresos.

De acuerdo a la modalidad del trabajo tenemos:

Ingreso monetario por Trabajo Dependiente

Comprende:

Ingreso líquido

Es el ingreso total monetario percibido por los trabajadores dependientes menos los descuentos de ley (seguro social, caja de pensiones, cesantía, jubilados, etc.) y otros descuentos (pago a cooperativas, bancos, asociaciones, judiciales, etc.).

Comprende: Ingreso Líquido permanente (remuneración básica y bonificaciones por decreto ley) y el Ingreso Líquido No Permanente (horas extras, refrigerio, movilidad, comisiones, etc.).

Ingreso total

Se refiere a todos los ingresos monetarios recibidos por trabajos realizados para un empleador o patrono.

Comprende: sueldos o salarios, ingresos por horas extras, bonificaciones, pago por concepto de refrigerio y movilidad, comisiones, etc.; antes de efectuar los descuentos de ley y otros descuentos. Estos pagos se reciben en forma regular o permanente.

Ingreso por Trabajo Independiente:

Comprende:

Ingreso Total

Comprende los ingresos monetarios por actividades industriales, comerciales o por actividades como profesional liberal o trabajador independiente. Se debe registrar el Ingreso Total, es decir, antes de deducir los gastos de operación del periodo de referencia mencionado y los pagos por impuestos a las ventas y servicios (IGV) y otros impuestos (impuesto a la renta, FONAVI).

- **Pago en especie**

Comprende los bienes y servicios recibidos por los trabajadores dependientes como remuneración o cualquier tipo de retribución. Incluye el valor estimado de las comidas, uniformes, servicios de salud, transporte y las viviendas cedidas gratuitamente en los centros de trabajo.

- **Ingresos extraordinarios por trabajo**

Esta referida a los Ingresos Extraordinarios por Trabajo, es decir, todos los pagos efectuados al Trabajador Dependiente en forma extraordinaria (no permanente) donde se incluyen las gratificaciones por Navidad, fiestas patrias, bonificaciones escolares, utilidades, participación en la comunidad laboral, etc. recibidos en los últimos 12 meses, por el trabajo actual en la ocupación principal o secundaria. Considere el ingreso total o bruto.

- **Otros ingresos extraordinarios**

Son los ingresos ocasionales por concepto de transferencias y otros ingresos recibidos por los miembros del hogar (no incluye los ingresos extraordinarios por trabajo).

Comprende los ingresos monetarios y/o en especie recibidos por los miembros del hogar de parte de instituciones, empresas o particulares, que no tienen un carácter regular o periódico y que no exigen contraprestación por parte de los miembros del hogar.

Comprende los ingresos por concepto de seguro de accidente o vejez, herencia, juegos de azar y otros ingresos ocasionales. Por excepción cuando el informante sea jubilado, se considerará su ingreso percibido por gratificación y escolaridad en los últimos 12 meses como Otros Ocasionales.

- **Transferencias corrientes**

Se considera a aquellos ingresos monetarios, así como también los que se recibe en especie con carácter regular sin contraprestación por parte de los miembros del hogar.

- **Rentas de la propiedad**

Esta referido a los ingresos monetarios, recibidos en los últimos 12 meses, por concepto de utilidades empresariales, intereses por depósitos en bancos, cooperativas; dividendos de acciones, bonos, arrendamiento de casas, departamentos, maquinarias y vehículos; arrendamiento de tierras agrícolas; alquiler de marcas, patentes, etc. Comprende:

- **Utilidades Empresariales:** Son los ingresos netos recibidos en el período de referencia por operaciones comerciales y financieras.
- **Intereses:** Son los ingresos netos recibidos en el período de referencia por los depósitos de ahorro, bonos, préstamos, etc.

- Dividendos: Son los ingresos netos recibidos en el período de referencia por concepto de utilidades de patrimonio accionario de una empresa que tiene el informante.
- Arrendamiento de Casas Departamentos, Habitaciones, Maquinarias y Vehículos: Son los ingresos netos recibidos en el período de referencia por el alquiler de bienes y muebles como edificios para oficinas y establecimientos, maquinaria y equipo.
- Arrendamiento de Tierras Agrícolas: Son los ingresos netos recibidos en el período de referencia por el uso de tierras agrícolas.
- Otro: Incluye alquiler de marcas, patentes, derechos de autor.

Metodología para la obtención, generación de variables calculadas a nivel de hogar - Sumaria:

Después de los procesos de imputación y deflactación se ejecuta el proceso de cálculo de la sumaria, el cual consiste en incorporar a la base de datos nuevas variables que son producto de agregaciones o combinaciones de preguntas. Se generan variables a nivel de hogar, las variables generadas dependen de los trimestres en curso, así para los trimestres I, II, III las variables son a nivel de ingresos, para el IV trimestre las variables son a nivel de gastos e

ingresos del hogar llevadas a un mismo período de investigación, para su posterior procesamiento y análisis.

Para el IV Trimestre del 2002, las variables generadas de gastos se agruparon a través de los ocho Grandes Grupos de Gastos y a la vez por sub grupos: (1) Gastos, (2) Autosuministro, pago en especie, (3) Donado, regalado; y (4) No gasto o no sabe. A continuación se detallan la relación de variables.

FACTORTO	: Factor de Expansión de Población
MIEPERHO	: Total de miembros del hogar
TOTMIEHO	: Total de personas en el hogar
PERCEPHO	: Total de perceptores de ingresos
INGREUHD	:Ingreso bruto por la act. principal dep. Monetario (dependiente)
INGNETHD	:Ingreso neto por la act. principal dep. Monetario (dependiente)
PAGESPHD	:Ingreso por pago en especie
INGPEDHD	:Ingreso permanente de la actividad principal (dependiente)
INGINDHD	: Ingreso por actividad principal independiente
INGAUTHD	: Ingreso por autoconsumo de la actividad principal
INGPEIHD	:Ingreso permanente de la actividad principal (independiente)
INSEILHD	:Ingreso neto por la actividad secundaria dependiente Monetario
PAESECHD	:Ingreso pago en especie por la actividad secundaria dependiente
INGSEIHD	:Ingreso neto por la act. Secundaria independiente Monetario
ISECAUHD	:Ingreso por autoconsumo de la act. Secundaria independiente
INGESTA	:Ingreso extraordinario por trabajo
INGTRAHD	:Ingreso por transferencias corrientes monetarias del país
INGTEXHD	:Ingreso por transferencias corrientes del extranjero
INGRENHD	:Ingreso por rentas de la propiedad monetario

INGOEXHD	:Otros Ingresos extraordinarios
IA01HD	:Ingreso alquiler imputado de la vivienda
IA02HD	:Ingreso por transferencia de alquiler
GA03HD	:Gasto por alquiler
GA04HD	:Gasto por alquiler imputado
G01HD	:Gastos por ampliación del hogar - Recursos Propios
IG02HD	:Ingresos y Gastos por ampliación del hogar - Donación, Regalo
IG03HD	:Ingresos y Gasto por ampliación del hogar - Crédito total
G04HD	:Gastos por amortización del Crédito
G05HD	:Gastos por alimentos consumidos fuera del hogar - Pagó
IG06HD	:Ingresos y Gastos por alimentos consumidos fuera del hogar - No Pagó
G07HD	:Otros Gastos - Pagó
IG08HD	:Ingresos y Gastos, Otros - No Pagó
GRU11HD	:Grupo 1: Alimentos – Gasto
GRU12HD1	:Grupo 1: Alimentos - Autosuministro, Autoconsumo
GRU12HD2	:Grupo 1: Alimentos - Pago en especie
GRU13HD1	:Grupo 1: Alimentos – Donación Público
GRU13HD2	:Grupo 1: Alimentos – Donación Privada
GRU13HD3	:Grupo 1: Alimentos - Donado, regalado, otro
GRU21HD	:Grupo 2: Vestido y Calzado – Gasto
GRU22HD1	:Grupo 2: Vestido y Calzado – Autoconsumo, Autosuministro
GRU22HD2	:Grupo 2: Vestido y Calzado - Pago en especie
GRU23HD1	:Grupo 2: Vestido y Calzado – Donación Pública
GRU23HD2	:Grupo 2: Vestido y Calzado – Donación Privada
GRU24HD	:Grupo 2: Vestido y Calzado - No gasto, no sabe
GRU31HD	:Grupo 3: Alquiler de Vivienda, Combustible, Electricidad y Conservación de la Vivienda – Gasto
GRU32HD1	:Grupo 3: Alquiler de Vivienda, Combustible, Electricidad y Conservación de la Vivienda – Autoconsumo, Autosuministro
GRU32HD2	:Grupo 3: Alquiler de Vivienda, Combustible, Electricidad y Conservación de la Vivienda - Pago en especie
GRU33HD1	:Grupo 3: Alquiler de Vivienda, Combustible, Electricidad y Conservación de la Vivienda – Donación Pública
GRU33HD2	:Grupo 3: Alquiler de Vivienda, Combustible, Electricidad y Conservación de la Vivienda – Donación Privada
GRU33HD3	:Grupo 3: Alquiler de Vivienda, Combustible, Electricidad y Conservación de la Vivienda - Otro

GRU34HD	:Grupo 3: Alquiler de Vivienda, Combustible, Electricidad y Conservación de la Vivienda - No gastó, no sabe
GRU41HD	:Grupo 4: Muebles y Enseres, y Mantenimiento de la Vivienda - Gasto
GRU42HD1	:Grupo 4: Muebles y Enseres, y Mantenimiento de la Vivienda – Autoconsumo - Autosuministro
GRU42HD2	:Grupo 4: Muebles y Enseres, y Mantenimiento de la Vivienda - Pago en especie
GRU43HD1	:Grupo 4: Muebles y Enseres, y Mantenimiento de la Vivienda – Donación Pública
GRU43HD2	:Grupo 4: Muebles y Enseres, y Mantenimiento de la Vivienda – Donación Privada
GRU43HD3	:Grupo 4: Muebles y Enseres, y Mantenimiento de la Vivienda - Otros
GRU44HD	:Grupo 4: Muebles y Enseres, y Mantenimiento de la Vivienda - No gastó, no sabe
GRU51HD	:Grupo 5: Cuidado, Conservación de la Salud y Servicios Médicos - Gasto
GRU52HD1	:Grupo 5: Cuidado, Conservación de la Salud y Servicios Médicos – Autosuministro, Autoconsumo
GRU53HD1	:Grupo 5: Cuidado, Conservación de la Salud y Servicios Médicos – Donación Pública
GRU53HD2	:Grupo 5: Cuidado, Conservación de la Salud y Servicios Médicos – Donación Privada
GRU53HD3	:Grupo 5: Cuidado, Conservación de la Salud y Servicios Médicos - Otros
GRU54HD	:Grupo 5: Cuidado, Conservación de la Salud y Servicios Médicos - No gastó, no sabe
GRU61HD	:Grupo 6: Transportes y Comunicaciones - Gasto
GRU62HD1	:Grupo 6: Transportes y Comunicaciones - Autosuministro, Autoconsumo
GRU62HD2	:Grupo 6: Transportes y Comunicaciones - Pago en especie
GRU63HD1	:Grupo 6: Transportes y Comunicaciones – Donación Pública
GRU63HD2	:Grupo 6: Transportes y Comunicaciones – Donación Privada
GRU63HD3	:Grupo 6: Transportes y Comunicaciones - Otro
GRU64HD	:Grupo 6: Transportes y Comunicaciones - No gastó, no sabe
GRU71HD	:Grupo 7: Esparcimiento, Diversión, Servicios Culturales y de Enseñanza - Gasto
GRU72HD1	:Grupo 7: Esparcimiento, Diversión, Servicios Culturales y de Enseñanza - Autosuministro, Autoconsumo
GRU72HD2	:Grupo 7: Esparcimiento, Diversión, Servicios Culturales y de Enseñanza - Pago en especie
GRU73HD1	:Grupo 7: Esparcimiento, Diversión, Servicios Culturales y de Enseñanza – Donación Pública

GRU73HD2	:Grupo 7: Esparcimiento, Diversión, Servicios Culturales y de Enseñanza – Donación Privada
GRU73HD3	:Grupo 7: Esparcimiento, Diversión, Servicios Culturales y de Enseñanza - Otros
GRU74HD	:Grupo 7: Esparcimiento, Diversión, Servicios Culturales y de Enseñanza - No gastó, no sabe
GRU81HD	:Grupo 8: Otros Bienes y Servicios - Gasto
GRU82HD1	:Grupo 8: Otros Bienes y Servicios - Autosuministro, Autoconsumo
GRU82HD2	:Grupo 8: Otros Bienes y Servicios - Pago en especie
GRU83HD1	:Grupo 8: Otros Bienes y Servicios – Donación Pública
GRU83HD2	:Grupo 8: Otros Bienes y Servicios – Donación Privada
GRU83HD3	:Grupo 8: Otros Bienes y Servicios - Otros
GRU84HD	:Grupo 8: Otros Bienes y Servicios - No gastó, no sabe
SG42	:Gastos por equipamiento del hogar
INGMO1HD	:Ingreso monetario (Bruto) “A” trimestral
INGMO2HD	:Ingreso monetario (Neto) “B” trimestral
INGHOG1D	:Ingreso total (Bruto) “A” trimestral
INGHOG2D	:Ingreso total (Neto) “B” trimestral
GASHOG1D	:Gasto Monetario Trimestral
GASHOG2D	:Gasto Total Trimestral

Las variables generadas de gastos se agrupan tal como se trabaja para el cálculo del Índice de Precios al Consumidor en ocho grandes grupos de gastos y para cada uno de estos grupos se distinguió la forma como se adquirió dicho bien o servicio. Así tenemos, que se ha distinguido si el bien es adquirido por compra; autoconsumo, autosuministro o pago en especie; donado por algún programa social o algún particular, y por último si no sabe o no efectuó gasto por alguna otra razón. A continuación se detallan la relación de variables que se incluye en el registro que contiene las variables agrupadas.

El Ingreso total del hogar se obtiene como producto de la suma de las variables siguientes:

$$\begin{aligned} \text{INGHOG1D} = & \text{INGBRUHD} + \text{INGINDHD} + \text{INSETHD} + \text{INGSEIHD} + \text{INGEXTHD} + \text{INGTRAHD} \\ & + \text{INGTEXHD} + \text{INGRENHD} + \text{INGOEXHD} + \text{PAGESPHD} + \text{INGAUTHD} + \text{PAESECHD} + \\ & \text{ISECAUHD} + \text{IA01HD} + \text{IG03HD} + \text{IG06HD} + \text{IG08HD} + \text{SIG24} + \text{GRU13HD} + \text{GRU23HD} + \\ & \text{GRU24HD} + \text{GRU33HD} + \text{GRU34HD} + \text{GRU43HD} \\ & + \text{GRU44HD} + \text{GRU53HD} + \text{GRU54HD} + \text{GRU63HD} + \text{GRU64HD} + \text{GRU73HD} + \text{GRU74HD} + \\ & \text{GRU83HD} + \text{GRU84HD} \end{aligned}$$

El Gasto total del hogar se obtiene como producto de la suma de las variables siguientes:

$$\begin{aligned} \text{GASHOG2D} = & \text{G05HD} + \text{IG06HD} + \text{G07HD} + \text{IG08HD} + \text{SG23} + \text{SIG24} + \text{SG42} + \text{GRU11HD} + \\ & \text{GRU12HD} + \text{GRU13HD} + \text{GRU21HD} + \text{GRU22HD} + \text{GRU23HD} + \text{GRU24HD} + \text{GRU31HD} + \\ & \text{GRU32HD} + \text{GRU33HD} + \text{GRU34HD} + \text{GRU41HD} + \text{GRU42HD} + \text{GRU43HD} + \text{GRU44HD} + \\ & \text{GRU51HD} + \text{GRU52HD} + \text{GRU53HD} + \text{GRU54HD} + \text{GRU61HD} + \text{GRU62HD} + \text{GRU63HD} + \\ & \text{GRU64HD} + \text{GRU71HD} + \text{GRU72HD} + \text{GRU73HD} + \text{GRU74HD} + \text{GRU81HD} + \text{GRU82HD} + \\ & \text{GRU83HD} + \text{GRU84HD} \end{aligned}$$

Período de referencia

Los períodos de referencia para las variables a investigarse en la Encuesta Nacional de Hogares 2003, son los siguientes:

Características de la Vivienda y del Hogar:

- Día de la entrevista
- Últimos 15 días
- Último mes
- Últimos 12 meses.

Características de los Miembros del Hogar:

- Día de la entrevista.
- Residencia habitual: Últimos 30 días.

Condición de Ocupación: Semana Pasada

Es la semana calendario anterior al día de la entrevista como tiempo fijado respecto al cual se investiga la condición de actividad.

Educación

- Día de la entrevista
- Último mes
- Presente año (2003)
- Últimos 12 meses

Salud

- Día de la entrevista
- Últimas 4 semanas
- Últimos 3 meses
- Últimos 12 meses

Empleo

- Condición de actividad: Semana pasada (semana calendario anterior al día de la entrevista)
- Ocupados y Desocupados: Semana pasada
- Total Horas
- Trabajo anterior para los desocupados e inactivos: Último trabajo realizado

Ingreso por Trabajo

- Ordinarios: Día, semana, quincena o mes anterior
- En Especie: Últimos 12 meses
- Extraordinarios Por Trabajo Dependiente: Últimos 12 meses
- Ingreso por Actividad Agropecuaria: Últimos 12 meses
- Ingreso del Trabajador Independiente o Empleador o Patrono: Mes anterior / Últimos 12 meses

Otros Ingresos

- Ingresos por Transferencias Corrientes: Últimos 6 meses
- Ingresos por Rentas de la Propiedad: Últimos 12 meses
- Otros Ingresos Extraordinarios: Últimos 12 meses
- Sistema de Pensiones: Día de la entrevista
- Programas Sociales: Últimos 12 meses

Gastos

- Últimos 15 días
- Mes anterior
- Últimos 3 meses
- Últimos 12 meses
- Últimos 10 años

Programas Sociales

- Últimos 3 meses
- Últimos 12 meses

VI. Arquitectura de las Bases de Datos**VI.1 Estructura y naturaleza de los archivos de datos**

Para la elaboración de archivos de datos finales. En general, se ha tomado como modelo la estructura que presentan los módulos en cada cuestionarios de la encuesta.

Veamos algunos ejemplos; en los resultados de la ENAHO 2002-IV Trimestre, tenemos:

Cuestionario	Modulo	Archivo SPSS
Enaho 01	100. Características de Vivienda y el Hogar	Enaho01-2002IV-100.SAV
	200. Características de los Miembros del Hogar	Enaho01-2002IV-200.SAV
	600. Gastos del Hogar – Gastos en alimentos y bebidas	Enaho01-2002IV-601.SAV
Enaho 01A	300. Educación	Enaho01A-2002IV-300.SAV
	400. Salud	Enaho01A-2002IV-400.SAV
	500. Empleo e Ingreso	Enaho01A-2002IV-500.SAV

En cuanto a la naturaleza de los archivos de datos, estos se clasifican en:

- Archivos del Hogar.
- Archivos de Personas o Miembros del Hogar.

VI.2 Identificación de casos y otras características

Dada la naturaleza distintas de los archivos de datos (Hogares, Personas) También corresponde una identificación distinta de ambos casos.

- Si el archivo es de Hogares la identificación es : Conglomerado + Vivienda + Hogar.
- Si el archivo es de Personas la identificación es : Conglomerado + Vivienda + Hogar + Persona.

VI.3 Coexistencia de variables

Otras de las características en los archivos de datos, es que podemos encontrar según sea el caso (Archivos de Hogares y/o de Personas) que:

- Las variables digitadas: P105B, P106 (variables 105 y 106 del cuestionario Enaho.01).
- Las variables deflactadas: D105B, D106 .
- Las variables imputadas: I105B, I106 .
- Las variables recodificadas, si las hubiera; T110, T211, etc.

Es decir, siempre se mantiene el valor original de cada variable.

VI.4 Generación de Variables Calculadas por Hogares

Para un mejor estudio de los ingresos y gastos efectuados en el hogar, se genera como archivo final el archivo de Variables Calculadas del Hogar o archivo Sumaria del Hogar.

Entre la variables que podemos encontrar tenemos:

- Factor de Expansión.
- Total de miembros del hogar.
- Total de Perceptores de Ingresos.
- Gastos del Hogar – Agrupados en 8 grandes grupos de gastos.
- Ingresos Monetario trimestral.
- Ingreso Total Trimestral.
- Gasto Monetario Trimestral.
- Gasto Total Trimestral.
- Línea de Pobreza.

VII. Generación de Tabulados

VII.1 Uso de Sistemas Integrados para el Procesamiento de Encuestas

Para cada uno de los proceso y/o subprocesos implementados en la encuesta existe un software utilizado, tal como podemos apreciar en el siguiente cuadro:

PROCESO	SUB-PROCESO	SOFTWARE BASE
Procesamiento Primario	Entrada de Datos	ISSA-X
	Cobertura	ISSA-X
	Codificación Automática	Visual FoxPro 6.0
	Consistencia Primaria	ISSA-X
Procesamiento Final	Consistencia Final	SPSS
	Deflactación	ISSA-X
	Imputación	ISSA-X
	Cálculo del Factor de Expansión	ISSA-X
Análisis de la Información	Errores Muestrales	STATA
	Generación de Tabulados	SPSS, STATA
	Cálculo de Indicadores (Línea de Pobreza, Informalidad, ..., etc.)	SPSS, STATA

Si bien es cierto que ISSA-X es un software para el uso en encuestas, tiene como inconveniente que necesita de un entorno integrador para sus aplicaciones. Por ello, es necesario el empleo del FoxPro 2.6 para la generación de un menú de opciones, que puedan invocar a cada una de las aplicaciones desarrolladas en ISSA-X.

En el procesamiento de la Encuesta Nacional de Hogares 2003, se está empleando VISUAL BASIC 6.0 para el desarrollo mixto de un módulo de procesamiento final de datos, es decir, hay tareas realizadas con aplicaciones ISSA-X y otros conjunto de tareas a realizarse directamente en VISUAL BASIC 6.0.

El menú de opciones de procesamiento está en Visual Basic y las aplicaciones que ejecutan tareas están en ISSA-X, aunque el objetivo siguiente será migrar de ISSA-X hacia CSPRO.



**DIFUSIÓN DE INFORMACIÓN DE LA ENCUESTA
NACIONAL DE HOGARES**

INSTITUTO NACIONAL DE ESTADÍSTICA E INFORMÁTICA (INEI)

PERÚ

ÍNDICE

		<u>Página</u>
I.	Política de Difusión Institucional	247
	I.1 Lineamientos de la política de difusión	247
	I.2 Impacto de la socialización de la información.....	247
	I.3 Política de ventas	247
II.	Sistema de Microdatos INEI.....	248
	II.1 Antecedentes.....	248
	II.2 Definición y objetivo.....	248
	II.3 Diagrama de flujo de datos.....	248
	II.4 Diagrama de estructura.....	249
	II.5 Características del sistema.....	249
	Disponibilidad de información organizada	249
	Datos en archivos formato estandar	250
	Acceso administrativo a la base de datos.....	250
	Atención personalizada en forma permanente	251
	II.6 Requerimientos mínimos de hardware y software.....	251
	II.7 Herramientas de desarrollo.....	251

I. Políticas de Difusión Institucional

I.1 Lineamientos de la Política de Difusión

Los lineamientos sobre los cuales se enmarca la Política de Difusión se sustentan en el Principio Básico de la Difusión, el cual señala que **“La Difusión de la información estadística debe ser amplia, oportuna, de fácil acceso, objetiva y coherente”** todo ello dentro de un contexto de modernización institucional que responda a satisfacer la, cada vez, más creciente demanda de datos estadísticos.

Se han trazado los grandes lineamientos siguientes:

- Realizar una amplia difusión de la información estadística que produce el Sistema Nacional de Estadística (SNE). Satisfacer las necesidades de información del Estado y de cada uno de los estamentos de la sociedad civil, a través de la socialización de la información.
- Diseñar y establecer una Línea Editorial, acorde con las necesidades de los usuarios que nos permita alcanzar una identidad editorial.
- Hacer uso intensivo de las nuevas tecnologías de la información, como vehículo de difusión, que nos permitan ampliar la cobertura y uso de la información optimizando los recursos.
- Adoptar procesos modernos de mercadotecnia en las actividades vinculadas a la comercialización de los productos y servicios que produce el INEI, que nos permita generar los recursos propios que retroalimenten la producción del INEI.

I.2 Impacto de la socialización de la información

- Con la disponibilidad de la información se ha logrado transparencia en la opinión especializada debido a que se publica toda la información básica (microdato) y también la documentación técnica que la sustenta.
- Aumento en la calidad de los datos, pues permite ir incorporando las recomendaciones y sugerencias de aquellas personas que cumplen roles preponderantes en la sociedad, lo que conlleva a mejorar la credibilidad de la institución.
- Creciente expectativa por la obtención de la información, también en el estado, lo cual está altamente relacionado con la obtención de los recursos presupuestales necesarios para su ejecución.
- Fomento de convenios con las instituciones nacionales e internacionales en la ejecución de proyectos similares.

I.3 Políticas de ventas

La política de ventas institucional esta definida por la necesidad de lograr una amplia difusión de la información producida, poniendo a disposición de los usuarios, públicos y privados, los diferentes productos y servicios que elabora el INEI, sin dejar de considerar nuestra condición de entidad del Estado, respetando las normas vigentes y particularmente cuidando el carácter de secreto estadístico de la información proporcionada por los informantes.

En este contexto el INEI, ha definido una política de costos asociado al tipo de información, relevancia y periodicidad de la misma, sin embargo comprometido con el apoyo a la comunidad, brinda en forma gratuita el acceso a la información básica de las encuestas que realiza, entre las que se encuentra la Encuesta Nacional de Hogares (ENAHOG). El usuario debe indicar la fuente de datos, si hiciera uso de ella para sus investigaciones.

II. Sistema de Microdatos INEI

II.1 Antecedentes

En el marco del Proyecto MECOVI-PERU, que ha respaldado ampliamente, todo lo referente a la ENAHO, se tomó como prioridad la difusión de la información proveniente de las encuestas de hogares, es así que el INEI decide publicarla en un medio masivo como la Internet y define que se proveerá en archivos cuyos formatos permitan un fácil manejo de los datos de la información básica (nivel microdato).

Para esto se desarrolló el denominado Sistema de Microdatos INEI, el cual alojado en el Web Site del INEI se lanza al público en el año 2000 y se mantiene operativo hasta la fecha con una gran aceptación por parte de los usuarios.

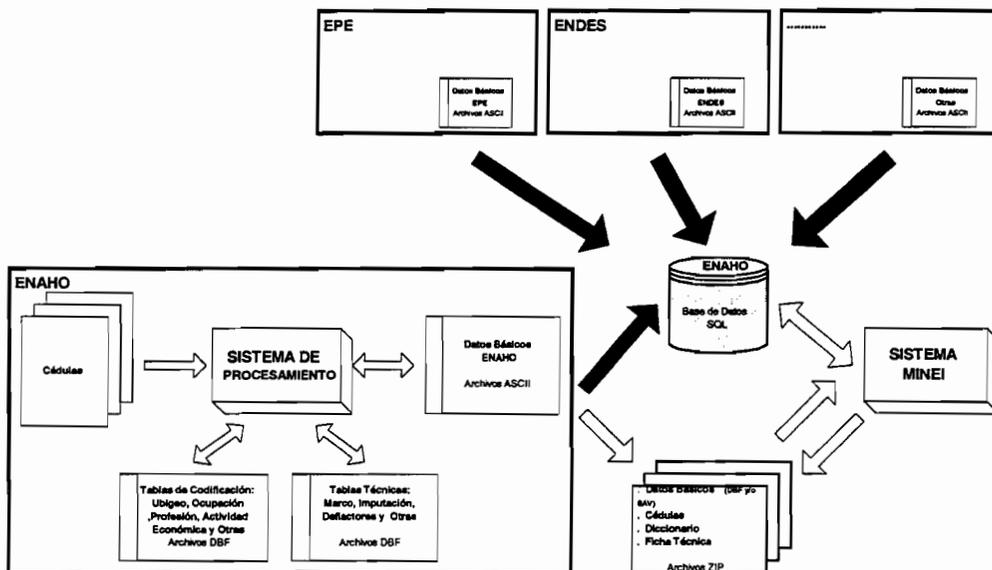
II.2 Definición y objetivo

El MINEI Ver. 1.0: **Sistema de Microdatos INEI**, es un sistema de difusión de información estadística proveniente de las encuestas realizadas en el INEI y ha sido desarrollado utilizando herramientas de última generación en plataforma Internet, considerando opciones dinámicas y transparentes para los diferentes tipos de usuarios involucrados en el sistema, como por ejemplo: descarga y solicitud de módulos de encuesta en los tipos de formato más usados por los investigadores y estadísticos, envío de e-mails automáticos que agilizan el mantenimiento de usuarios, entre otros.

Entre las encuestas de las cuales se publican datos en el MINEI tenemos: La Encuesta Nacional de Hogares (ENAHO), la Encuesta Permanente de Empleo (EPE), la Encuesta Nacional de Demografía y Salud (ENDES) y otras.

II.3 Diagrama de flujo de datos

FLUJO DE DATOS DEL SISTEMA MINEI



II.4 Diagrama de estructura



II.5 Características del sistema

Disponibilidad de información organizada

En el MINEI a partir de mayo del 2001, se publican datos de diferentes encuestas, incluido un registro de estadísticas vitales, las encuestas están organizadas en temas o módulos y se tiene información de 1994 a la fecha, cabe mencionar que más del 80% de módulos corresponden a la Encuesta Nacional de Hogares (ENAH).

NUMERO DE MODULOS Y PERIODOS POR ENCUESTA

Encuesta	Periodo (Años)	Número de Módulos	
		Absoluto	%
TOTAL		375	100.00
ENAH	1997 - 2001	312	83.20
ENDES	1996 y 2000	20	5.33
EPE	2001 - 2003	25	6.67
EV-Nacimientos	1999 - 2000	2	0.53
ENAPROM	1994	16	4.27

Datos en archivos de formato estandar

A fin de facilitar el trabajo de los usuarios, quienes permanentemente requerían la conversión de los archivos a dichos formatos, el sistema permite recuperar los archivos en formato *.DBF y *.SAV (del SPSS), lo cuales pueden ser descargados directamente si corresponde a un solo módulo o ser solicitados si se trata de varios módulos en simultáneo.

Para una mayor agilidad en la descarga, los archivos de todas las encuestas se han organizado en archivos empaquetados y cada uno de ellos contiene el archivo de datos de un módulo (con extensión en DBF o SAV), el diccionario de datos, la ficha metodológica y la cédula escaneada.

Actualmente, el volumen total de datos de la ENAHO que el sistema maneja es de aproximadamente 572 MB, de lo cuales 93 MB corresponden a las tablas propias del sistema en la BD Sql y 479 MB a los archivos de datos básicos.

ENAHO- VOLUMEN DE DATOS POR AÑO

Año	Volumen (MB)		
	Tablas del Sistema	Datos en DBF	Datos en SAV
TOTAL	93	229	250
1997		42	25
1998		51	55
1999		95	128
2000		35	36
2001		4	4
2002		2	2

Acceso administrado a la base de datos

El MINEI permite el acceso gratuito a la información, mediante el uso de un código de usuario y una clave, la cual tiene periodo de caducidad al trimestre de haberlo concedido. El usuario puede solicitar la ampliación para su permanente acceso.

Los accesos se han determinado de acuerdo al tipo de usuario, los cuales tienen roles bien definidos:

Administrador

Encargado de la administración y mantenimiento de los usuarios del Sistema. Así mismo será el encargado de la actualización de la información de las Encuestas del INEI a ser publicada en el Web.

Validador

Responsable de evaluar los datos ingresados por los candidatos a Usuarios del sistema (usuarios registrados), cambiándoles su estado de Solicitante a Activo. Este Rol puede ser cumplida por el Administrador del Sistema o por otro Usuario con los permisos correspondientes. Toma conocimiento vía e-mail de un nuevo registro de usuario por cada evento realizado en el proceso de registro de usuario.

Usuario

A quien es dirigido el sistema (end user) y puede ver y descargar información de cualquier módulo de encuesta. Este usuario ingresará al sistema mediante una Cuenta de usuario y Contraseña que será validado por el usuario Validador (estos dos datos serán ingresados por el usuario en el módulo de "Registro de Usuarios" y serán refrendados por el Administrador). Los usuarios a los que está dirigido el

sistema están constituidos por investigadores, estudiantes, analistas estadísticos, etc. Toma conocimiento de que puede usar el sistema a través de un e-mail, enviado por el sistema.

Cabe mencionar que desde que se publicó el sistema en Internet, se han registrado 8686 usuarios.

NUMERO DE USUARIOS REGISTRADOS

Tipo de Usuario	Número de Usuarios				
	Total	Activos	%	Inactivos	%
TOTAL	8686	8148	100		100
Estudiante		1630	20.00		20.00
Empleado		6518	80.00		80.00
Investigador		6368	78.15		78.15
Trabajador independiente		77	0.95		0.95
Otro		73	0.90		0.90

Atención personalizada en forma permanente

Los requerimientos por parte de los usuarios se atienden en forma permanente, y están sujetos a registros, a fin de poder tomar decisiones respecto a políticas de difusión. Las estadísticas que nos proporciona el sistema son las siguientes:

ATENCIONES REALIZADAS A USUARIOS

Módulos por Proceso	Número de Atenciones		
	Todas las Encuestas	ENAH0	
		Absoluto	Absoluto
TOTAL		7600	
Descargados		1082	00.00
Solicitados	6518	6518	00.00
Atendidos	6368	6368	00.00
Pendientes	77	77	00.00
Descartados	73	73	00.00

II.6 Requerimientos mínimos de hardware y software

Equipo Cliente	Equipo Servidor Web
Hardware	
<ul style="list-style-type: none"> • Mainboard con procesador Intel Pentium 166 Mhz. • 32 Megabytes de memoria RAM. • Video con resolución de 800 x 600 • Monitor Super VGA de 14". • Conexión a Internet (usuarios) o Intranet (administradores). 	<ul style="list-style-type: none"> • Mainboard con 2 procesadores Intel Pentium III de 500 MHZ • 512 MB de RAM • Disco duro SCSI con capacidad de 10 Gigabytes. • Video con resolución de 800 x 600 • Monitor Super VGA de 14"

Equipo Cliente	Equipo Servidor Web
<ul style="list-style-type: none"> • Teclado • Mouse. 	<ul style="list-style-type: none"> • Tarjeta de Red • Teclado • Mouse.
Software	
<ul style="list-style-type: none"> • Sistema Operativo (Windows 95, Windows NT 4.0 Workstation, Windows ME, Windows 2000 Professional). • Navegador Internet Explorer 5.0 	<ul style="list-style-type: none"> • Sistema Operativo Windows NT Server 4.0 con Option Pack 4 y Service Pack 6 • Internet Information Server 4.0 con Service Pack 2 • Servicio SMTP del IIS 4.0 • MS SQL Server 7.0

II.7 Herramientas de desarrollo

Equipo Cliente	Equipo Servidor Web
Sistema Operativo	
<ul style="list-style-type: none"> • Windows 2000 Professional 	<ul style="list-style-type: none"> • Sistema Operativo: Windows NT Server con Option Pack 4 (como mínimo)
Software de Desarrollo	
<p>Herramientas de Programación</p> <ul style="list-style-type: none"> • Microsoft Visual Studio 6.0 (Visual Interdev 6.0 y Visual Basic 6.0). • Sistema de Gestión de Base de Datos: Microsoft SQL Server 7.0 Cliente <p>Tecnología empleada</p> <ul style="list-style-type: none"> • Tecnología Hipertext Markup Language (HTML), DHTML, ASP • VB Script y Java Script para la programación en el Cliente. • ADO para la comunicación con la fuente de datos (la base de datos en SQL Server). <p>Herramientas de Escritorio</p> <ul style="list-style-type: none"> • MS Office 2000 (MS Word con corrector ortográfico y gramática, MS Excel y MS Power Point) • MS Project 2000 • Visio 5.0 • Rational Rose 5.0 (ADOO - Notación UML). <p>Herramientas de Diseño</p> <ul style="list-style-type: none"> • Microsoft Front Page 97 • Adobe PhotoShop 5.5 • Corel Draw 	<p>Software de Gestión</p> <ul style="list-style-type: none"> • Sistema de Gestión de Base de Datos: Microsoft SQL Server 7.0 • Internet Information Server 4.0 con Service Pack 2 y con el servicio SMTP.

METODOLOGÍA PARA LA INTEGRACIÓN DE BASES DE DATOS DE ENCUESTAS DE HOGARES

BANCO CENTRAL

REPÚBLICA DOMINICANA



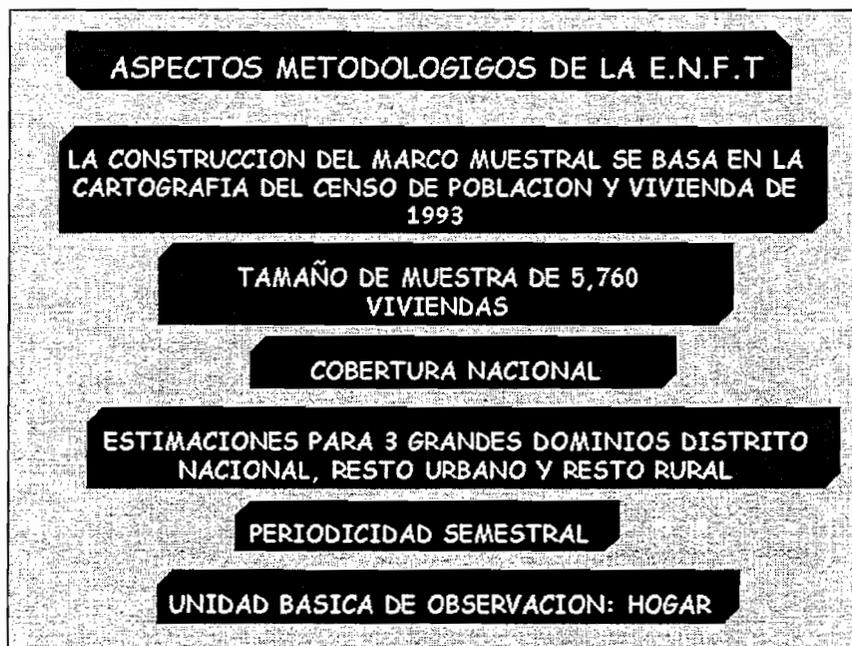
**BANCO CENTRAL DE LA
REPUBLICA DOMINICANA
(BCRD)**

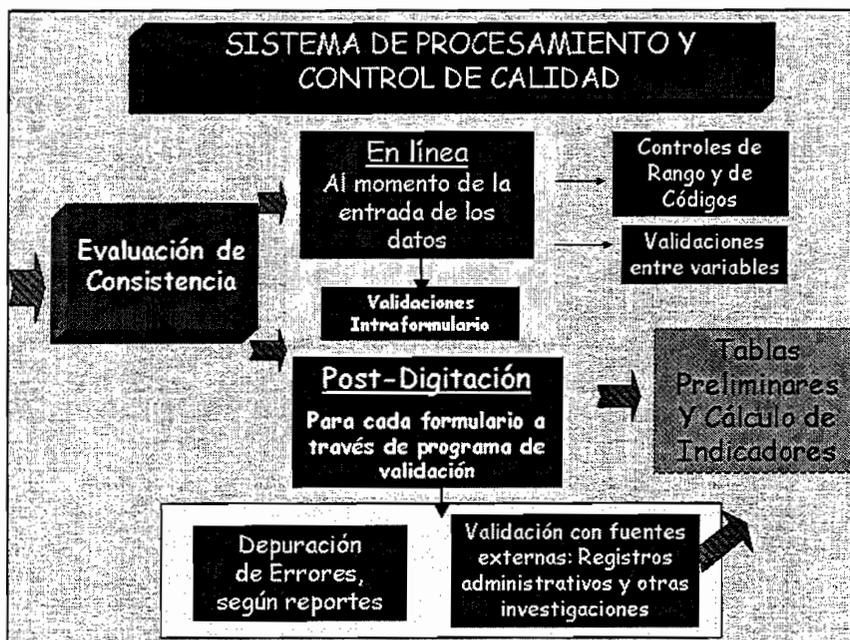
**Metodologías para la Integración
De bases de datos de
Encuestas de hogares**

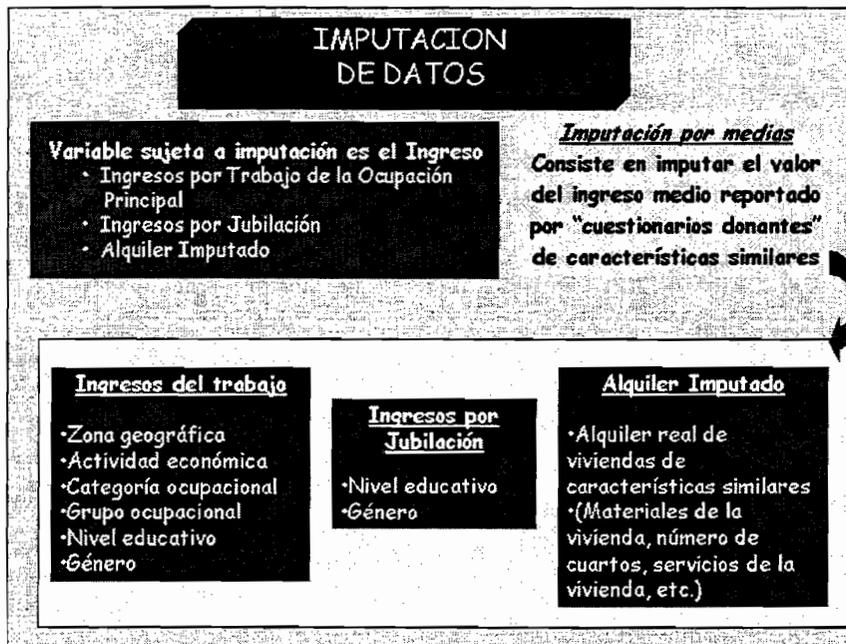
**TALLER REGIONAL CEPAL-INEI
Del 16 al 18 de Julio Lima, Perú**

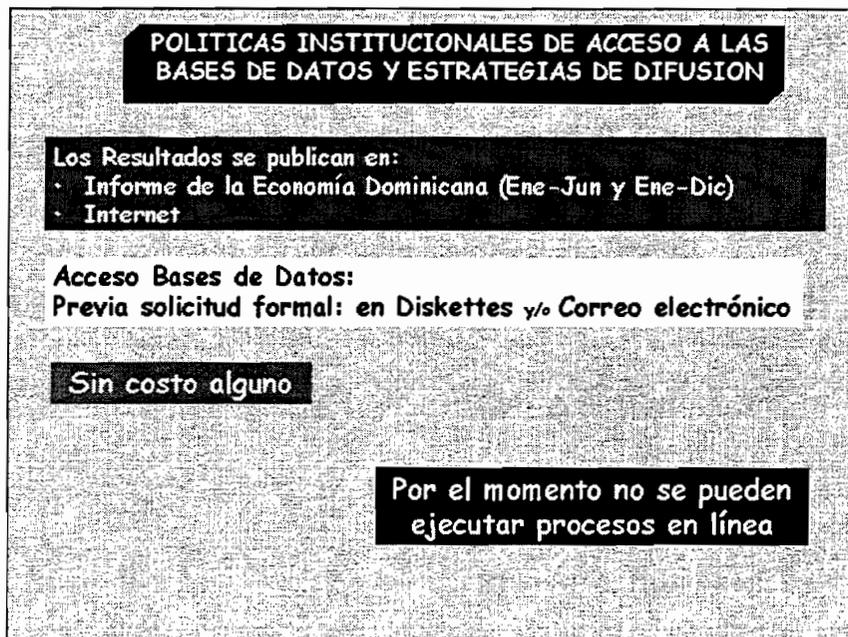
Encuestas de Hogares

Encuesta Nacional de Gastos e Ingresos de los Hogares	➔	Cada 10 años
Encuesta Nacional de Fuerza de Trabajo (ENFT)	➔	Semestral









HACIA DONDE VAMOS?

Nuevo marco muestral basado en la cartografía del Censo Nacional de Población de Octubre del 2002

Ampliación de la muestra como parte del mejoramiento de las encuestas de hogares
Para obtener resultados a niveles más desagregados

Programa MECOVI en la República Dominicana:

Capacitación en paquetes estadísticos y cartografía digitalizada

Explotación de los resultados de la ENFT

Generación de nuevos indicadores

**GRACIAS
POR SU ATENCION**

**METODOLOGÍAS PARA LA INTEGRACIÓN DE BASES
DE DATOS DE ENCUESTAS DE HOGARES**

CARLOS E. CALVO

INSTITUTO NACIONAL DE ESTADÍSTICA (INE)

URUGUAY

ÍNDICE

	<u>Página</u>
1. Introducción.....	263
2. Procedimientos de la ECH.....	263
3. Breve Descripción del Sistema ECH.....	264

El presente documento resume los aspectos principales del manejo de la información de la Encuesta Continua de Hogares (ECH) del Uruguay. Al considerar más conveniente e ilustrativo, realizar en el Taller, una demostración real del sistema utilizado por la Encuesta, solo se hace aquí una breve reseña del mismo. Por otra parte es posible acceder a la documentación completa de la Encuesta en el sitio web WWW.INE.GUB.UY¹; así como en intranet del Instituto se encuentran todos los manuales de procedimiento; en el marco de la definición del sistema de gestión de calidad del INE.

1. Introducción

Es importante, a los efectos de ilustrar la capacidad de este sistema, realizar una breve descripción de la ECH.

En el Uruguay desde el año 1968 se releva en forma ininterrumpida una encuesta a hogares, de propósitos múltiples (si bien sus objetivos básicos son el aportar información sobre la condición de actividad de la población y el nivel de ingresos familiar), de carácter continua (es decir se recoge información durante los 365 días del año), que divulga resultados mensualmente bajo la modalidad de trimestre móvil.

A pesar de que su ámbito geográfico es solo urbano, dado la distribución espacial de la población uruguaya, la información obtenida por la Encuesta Continua de Hogares (ECH) es representativa del 81 por ciento de la población total del país, tal cual se muestra a continuación:

<i>Población total proyectada año 2002 del Uruguay</i>	3.384.700
<i>Población proyectada año 2002 en localidades de 5000 o + habitantes</i>	2.731.194

El tamaño efectivo anual de la muestra es de 18.500 viviendas y 56.300 personas (cada persona entrevistada representó en el año 2002 a 48 habitantes).

El personal afectado a la Encuesta suma 56 personas distribuidas de la siguiente manera:

<i>Preparación de material y control administrativo de la encuesta</i>	5 funcionarios
<i>Relevamiento capital nacional</i>	12 funcionarios
<i>Relevamiento resto del país</i>	21 funcionarios
<i>Supervisión de campo</i>	7 funcionarios
<i>Consistencia, codificación e ingreso de encuestas</i>	8 funcionarios
<i>Análisis y generación de información</i>	3 funcionarios

Como se ve, el 70 por ciento del mismo se destina a la recolección en campo de la información y tan solo el 30 por ciento a todo el resto de las tareas incluyendo también dentro de éstas la elaboración y análisis de resultados.

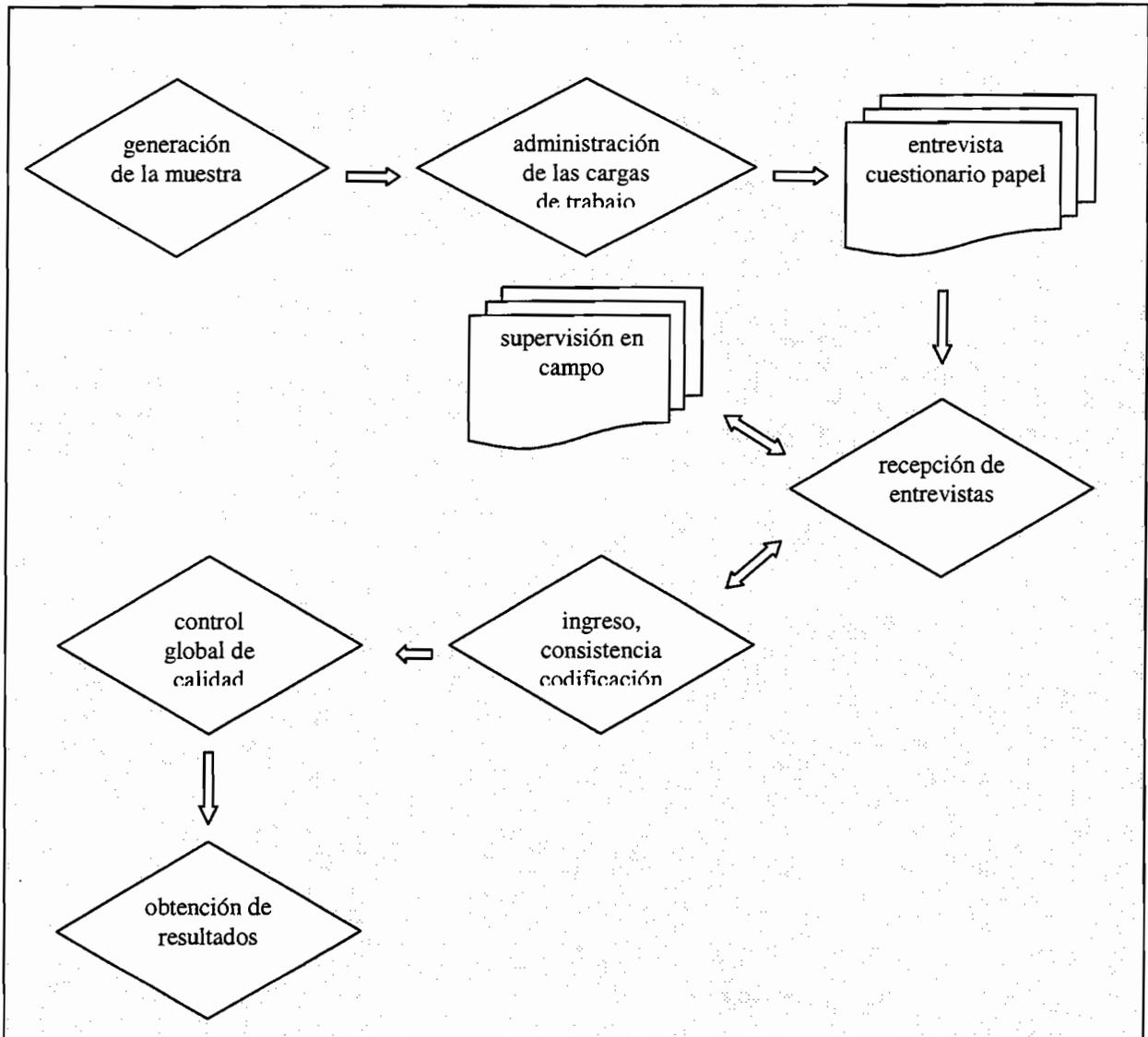
La difusión de la información, de acuerdo con el cronograma general del INE, se hace a los 30 días de finalizado el período de referencia.

2. Procedimientos de la ECH.

La Encuesta Continua de Hogares trabaja, a partir de la reformulación 2001 de su cuestionario, con un sistema integrando las actividades de relevamiento, administración, consistenciación, codificación y obtención de resultados básicos, lo que permitió mejorar la operativa general del trabajo en todas sus etapas, así como dar una total transparencia al mismo. La mejora se vio especialmente reflejada además, en un aumento sustantivo de la calidad ya que en su confección se puso el énfasis en las normas de control.

¹ "Biblioteca en línea, Series Monográficas, Metodologías, Metodología de la ECH".

Esquemáticamente el flujo de trabajo se muestra en la siguiente figura, donde las tareas encerradas en rombos se encuentran programadas en el sistema.



3. Breve Descripción del Sistema ECH

Tal como se indicó anteriormente el sistema ECH abarca las tareas desde la generación de la muestra hasta la tabulación de resultados; por lo que hace una integración total de las acciones.

La generación de la muestra se hace a partir del marco de viviendas del Censo General de Población y Viviendas. Esta operación se encuentra programada atendiendo a las especificaciones definidas de diseño, tamaño, confiabilidad, etc., y se dispara automáticamente.

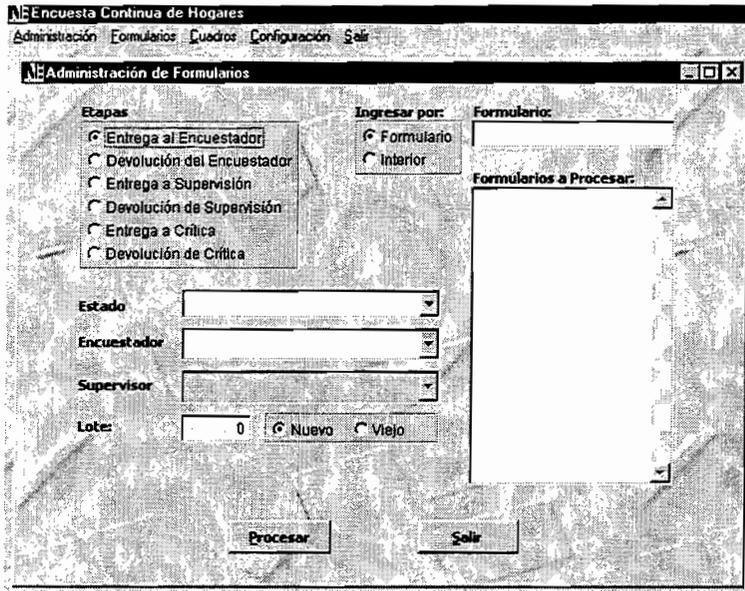


Resumidamente, el sistema se conecta con la base ORACLE, que contiene el marco del Censo (actualizado permanentemente por la Unidad de Cartografía), y de acuerdo con las especificaciones definidas, selecciona zonas censales dentro de las localidades participantes, y dentro de estas viviendas. En esta operación además se conecta con la base que contiene la muestra del año anterior, comparándola a efectos de que no exista duplicaciones. También calcula para cada vivienda su expansor mensual, anual y semestral, así como determina el mes y semana en que se debe relevar la misma.

A partir de esa muestra se ubican espacialmente las viviendas a encuestar (esta tarea es la única manual ya que los directorios de edificios detallados con que cuenta el Instituto actualmente no están en soporte magnético²) y automáticamente se genera e imprime una etiqueta autoadhesiva (que se pega a cada cuestionario) con los datos de identificación desplegados y en código de barras que permitirá realizar el seguimiento particular de cada cuestionario, es decir que funcionarios realizaron las distintas etapas, en que fechas, situación del mismo, etc.

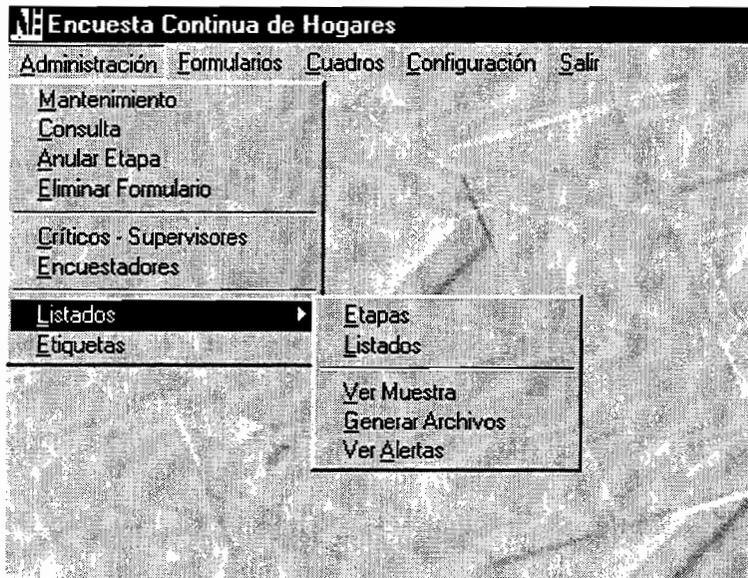
La asignación de cuestionarios a distintas tareas y funcionarios, así como la administración de la encuesta ocupa poco tiempo gracias a la utilización del código de barras. La siguiente figura muestra las pantallas para tales tareas.

² ya está el proyecto de digitación del directorio de edificios que surja del nuevo Censo.

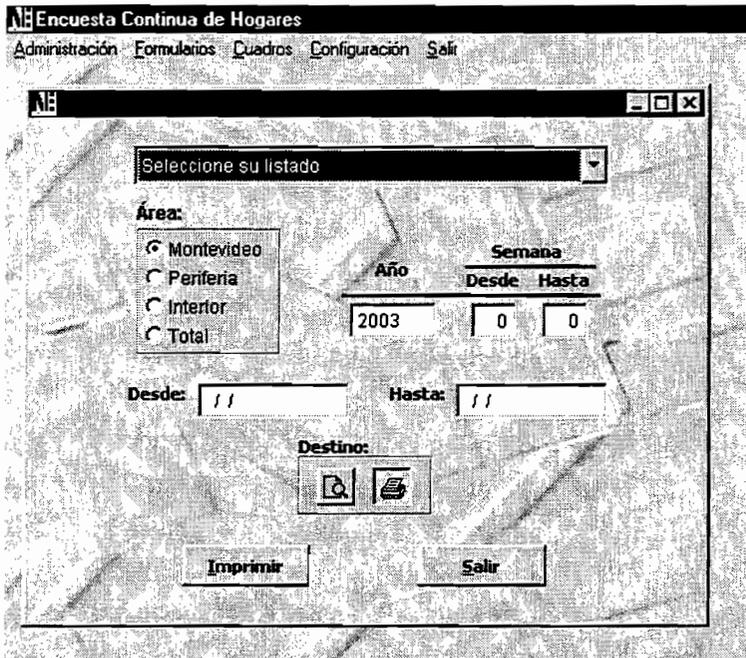


pantalla: ADMINISTRACION → MANTENIMIENTO

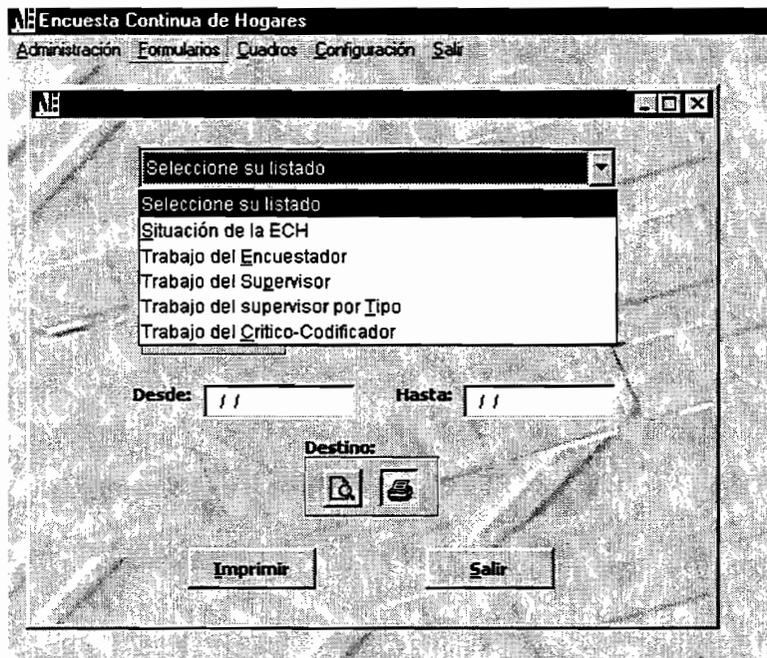
El sistema cuenta con una serie de reportes administrativos que permiten conocer en tiempo real la marcha del trabajo, así como efectuar la consulta en particular de cualquier entrevista.



La generación de cada uno de ellos puede realizarse para un período de tiempo en particular, así como para un área geográfica determinada.



pantalla: ADMINISTRACIÓN → LISTADOS → LISTADOS



pantalla: ADMINISTRACIÓN → LISTADOS → LISTADOS → Seleccione su listado

Estos reportes, se complementan con listados que detallan entrevista por entrevista, generados a través de la pantalla que muestra la figura siguiente. Por ejemplo con ellos se puede conocer el detalle de entrevistas que integra cada lote, etc.

Pantalla: ADMINISTRACION → LISTADOS → ETAPAS

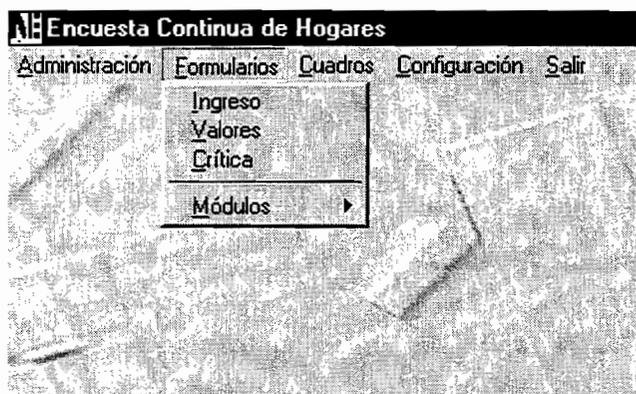
Mediante la aplicación VER MUESTRA, se puede realizar un seguimiento de todas las entrevistas en todas sus etapas, para analizar tiempos de trabajo, re-salidas a campo, etc. Es muy utilizada para ubicar entrevistas que puedan haber quedado sin digitar por traspapelarse en un lote. Como los listados se pueden generar en excel y para un período en particular, el manejo resulta muy simple.

Año	Semana	
	Desde	Hasta
2003	0	0

Pantalla: ADMINISTRACION → LISTADOS → VER MUESTRA

Esta aplicación puede realizarse para una entrevista en particular, utilizando la opción CONSULTA del sub-menú ADMINISTRACIÓN.

Una vez relevada la información, los cuestionarios son ingresados por un crítico-codificador mediante un programa inteligente que cuenta con un importante número (superior a 150) de reglas de consistencia (errores y alertas) que impide el ingreso de información con errores, omisiones, etc. *Es muy importante aclarar que la ECH en Uruguay no permite la omisión en ninguna de sus preguntas por tanto no hay imputación alguna.* En otro orden merece destacar, que el tiempo promedio de ingreso de toda la información de una persona es de 2 minutos y medio.



En esta misma operación se codifica la información de rama de actividad, tipo de ocupación, área geográfica, y educación. A tales efectos se cuenta con una base de datos de alimentación continua que permite codificar automáticamente o en su defecto en forma asistida la información.

Esta base de datos de codificadores no solo es utilizada por la ECH sino por todas las investigaciones que realiza el Instituto y necesiten codificar información de rama, ocupación, etc. Se debe destacar que esta base de datos cuenta con el nomenclator oficial de cada uno de los codificadores internacionales y de un diccionario (carteles o descripciones habituales) que se va incrementando permanentemente a través de la digitación de las descripciones anotadas por el encuestador, que quedan asociadas a los códigos asignados y controlados con el nomenclator oficial.

Además también en esta operación se valora, con parámetros estándar (los que deben cargarse, a través de la opción valores, todos los meses antes del ingreso del primer cuestionario para garantizar su actualización), ingresos en especie y se generan variables agregadas, tales como condición de actividad de la persona, inserción laboral (subempleo, registrado en la seguridad social, etc.), nivel de educación, ingreso de la ocupación principal, ingreso del hogar, etc., etc. variables que quedan adosadas al registro de cada persona u hogar según corresponda.

Algunas de las variables agregadas se presenta a continuación:

a. variables agregadas en registro de hogar.

⇒ cantidad de hombres.

⇒ cantidad de mujeres.

⇒ cantidad de menores de 14 años.

- ⇒ cantidad de mayores de 60 años.
- ⇒ cantidad de ocupados.
- ⇒ cantidad de desocupados.
- ⇒ cantidad de desocupados propiamente dichos.
- ⇒ cantidad de desocupados buscan trabajo primera vez.
- ⇒ cantidad de inactivos.
- ⇒ cantidad servicio doméstico.
- ⇒ cantidad de personas sin servicio doméstico.
- ⇒ cantidad de personas perceptores de ingresos.

b. variables de ingresos agregadas en el registro de hogar.

- ⇒ ingreso del hogar con valor locativo sin servicio doméstico.
- ⇒ ingreso del servicio doméstico.
- ⇒ valor locativo.
- ⇒ valor del alquiler.
- ⇒ ingreso del jefe de hogar.
- ⇒ ingreso del cónyuge.
- ⇒ ingreso otros miembros del hogar.

c. variables de ingresos agregadas en el registro de personas.

- ⇒ total de ingresos.
- ⇒ total de ingresos por trabajo
- ⇒ ingreso de la ocupación principal
- ⇒ ingreso de otras ocupaciones

d. determinación de la condición de actividad económica de la persona.

e. determinación de grupos especiales de ocupados.

f. determinación del nivel educativo de la persona

Una vez ingresadas todas las encuestas de un período determinado, el sistema cuenta con un módulo de crítica que realiza una consistenciación global de la base de datos seleccionada, emitiendo los reportes respectivos. Este procedimiento se completa con una serie de controles que se encuentran programados en SPSS, es decir por fuera del sistema, para extremar las medidas de calidad de la información, que exige, si es necesario la vista particular de entrevistas que se alejen de determinados patrones.

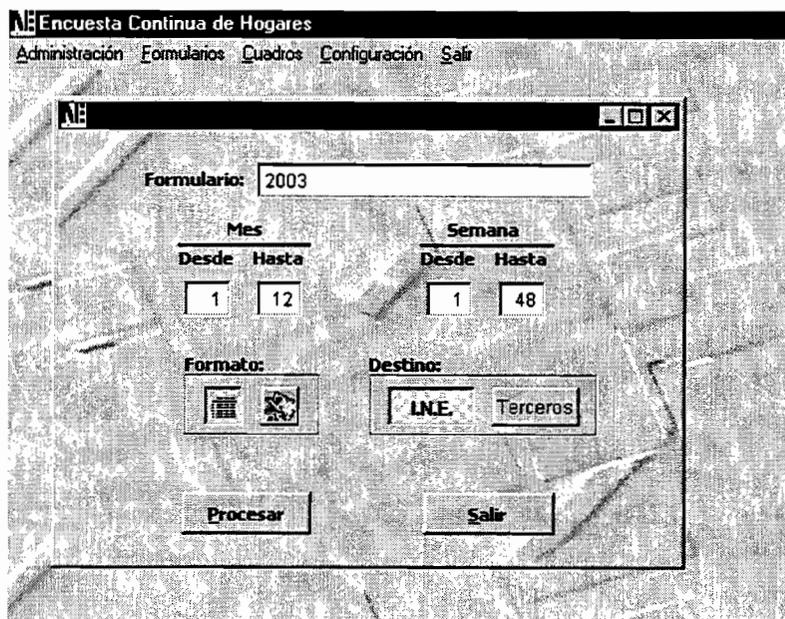
Aceptada la calidad de la base el sistema tiene programado un conjunto de cuadros que brinda la información, en planillas Excel, que de forma habitual difunde el Instituto sobre mercado de trabajo e ingresos familiares.

La nómina de estos cuadros se presenta a continuación:

1. Tasa de actividad, empleo y desempleo, por sexo, según área geográfica.
2. Distribución de la población ocupada, por sector de actividad, según área geográfica.
3. Distribución de la población ocupada, por sector de actividad, según área geográfica.
4. Promedio de horas trabajadas en la semana, por sector de actividad, según área geográfica.
5. Población ocupada por cantidad de empleos (%) y número medio de empleos por persona ocupada, según área geográfica.

6. Contribución a la tasa de desempleo, por componentes, sexo, edad y relación con la jefatura del hogar, según área geográfica.
7. Tasa específica de desempleo por grupo de edades, sexo y relación con la jefatura del hogar, según área geográfica.
8. Distribución del desempleo, por sector de actividad, según área geográfica.
9. Estimaciones de algunas medidas descriptivas del ingreso (en pesos corrientes) y la población, según área geográfica.
10. Distribución porcentual de desocupados, según área geográfica.
11. Tasa de actividad, empleo y desempleo, e ingreso medio del hogar, según gran área geográfica.
12. Personas desocupadas por condiciones de empleo requeridas (en porcentaje) y duración media del desempleo, según área geográfica.
13. Ingreso promedio (en pesos uruguayos) de la ocupación principal por categoría de la ocupación, según área geográfica y sexo.
14. Ingreso promedio (en pesos uruguayos) de la ocupación principal por sector de actividad, según área geográfica y sexo.
15. Distribución de la población ocupada por nivel más alto de educación alcanzado, según área geográfica.

Adicionalmente es posible la obtención de bases de microdatos para determinada área geográfica y período de tiempo, en formato Excel o DBF para trabajar en este último caso en SPSS y realizar todas las tablas a medida necesarias, en breves plazos.



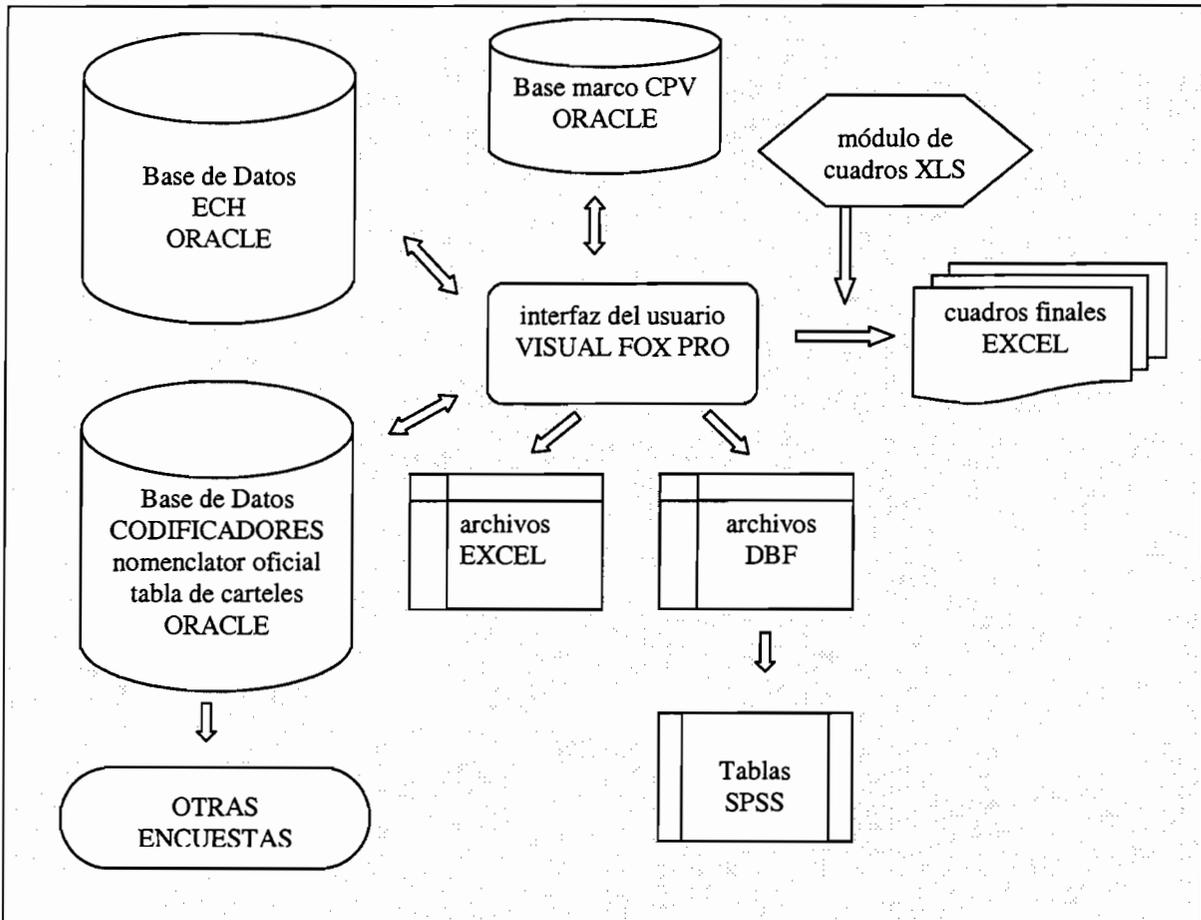
Pantalla: ADMINISTRACION → LISTADOS → GENERAR ARCHIVOS

Como se puede ver en la figura anterior, la generación de bases de microdatos contempla el destino de la misma, INE o Terceros. La diferencia fundamental que presentan estas es sobre la identificación de

registros para salvaguardar el secreto estadístico, así como inclusión de carteles o algunas variables generadas con el solo fin de control interno.

Como se desprende de lo anterior los usuarios tienen libre acceso a las bases de microdatos de la encuesta, lo que les permite realizar los cruces u obtener la información a su medida. Simplemente deben comprometerse a utilizarlas con los fines y en las condiciones que ellas permiten.

La figura siguiente intenta ilustrar el sistema ECH:



METODOLOGÍA PARA LA INTEGRACIÓN DE BASES DE DATOS DE ENCUESTAS DE HOGARES

INSTITUTO NACIONAL DE ESTADÍSTICA (INE)

VENEZUELA

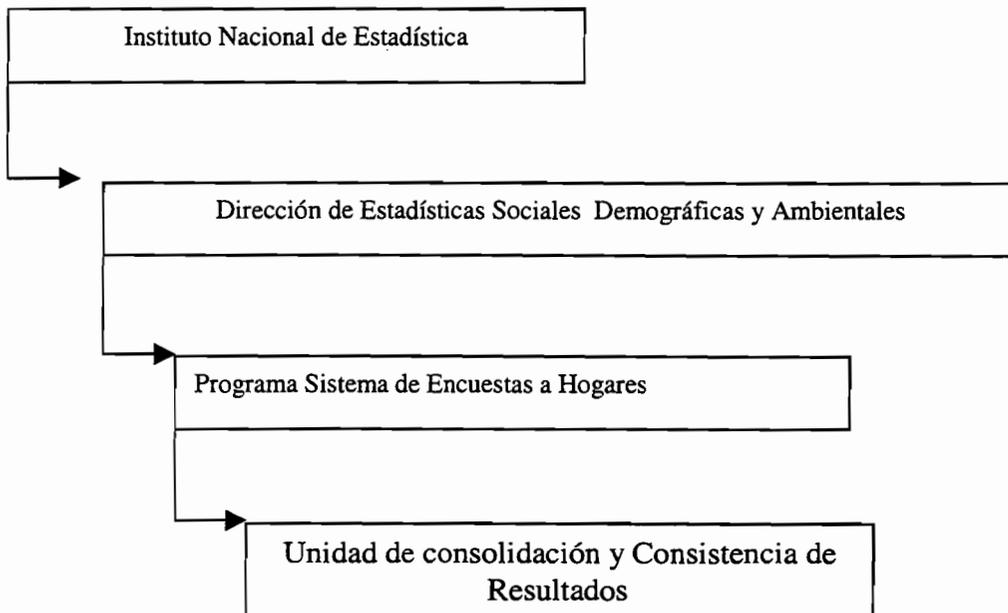
ÍNDICE

		<u>Página</u>
1.	Aspectos Institucionales	275
2.	Aspectos Generales.....	275
	Encuesta de hogares por muestreo	276
3.	Aspectos Computacionales de la EHM	277
	3.1 Estrategia para la captura de datos	277
	3.1.1 Centralización de los procesos	277
	3.1.2 Crítica y codificación de la planilla	278
	3.1.3 Captura de datos	279
	3.1.3.1 Diseño de estructura de datos.....	279
	3.1.3.2 Desarrollo de la aplicación.....	280
	3.1.3.3 Requerimiento de hardware y software.....	283
	3.1.3.4 Justificación para el desarrollo del sistema de captura	283
	3.2 Procedimientos para evaluar la consistencia de la información	283
	3.2.1 Validación	283
	3.3 Procesamiento de las bases de datos	284
	3.3.1 Software utilizados y desarrollo del mismo	284
	3.3.2 Cálculo de los factores de expansión	284
	3.3.3 Generación de tabulados	284
	3.4 Imputación de datos	285
	3.5 Integración de variables	285
	3.6 Integración de archivos de datos	285
	3.7 Políticas institucionales de acceso a la base de datos	286

1. Aspectos Institucionales

La Dirección de Estadísticas Sociales y Ambientales, dentro del INE, tiene como misión, garantizar que se efectúen los diseños de investigaciones estadísticas en el área social, demográfica y ambiental, así como, las proyecciones, predicciones y análisis estadístico que corresponda al INE y velar porque los diseños de las investigaciones que se realizan en otros órganos del Sistema Estadístico Nacional se efectúe en concordancia con el Plan Estadístico de la nación.

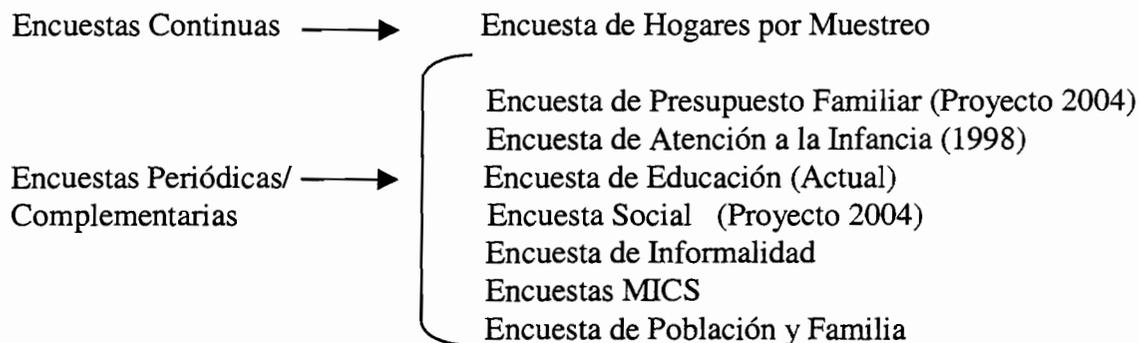
El Programa Sistema de Encuestas a Hogares, adscrito a la DESDA, es el encargado de velar por la armonización conceptual y metodológica de las Encuestas a Hogares del Sistema Estadístico Nacional, mediante la acción coordinada con las distintas instancias implicadas; en este programa se ubica la Unidad de Consolidación y Consistencia de Resultados, encargada de garantizar la ejecución de la fase relativas a la integración, consolidación y procesamiento de los datos de cada una de las encuestas del Sistema.



2. Aspectos Generales

Las Encuestas que integran actualmente el Programa de Sistemas de Encuestas a Hogares las podemos clasificar en:

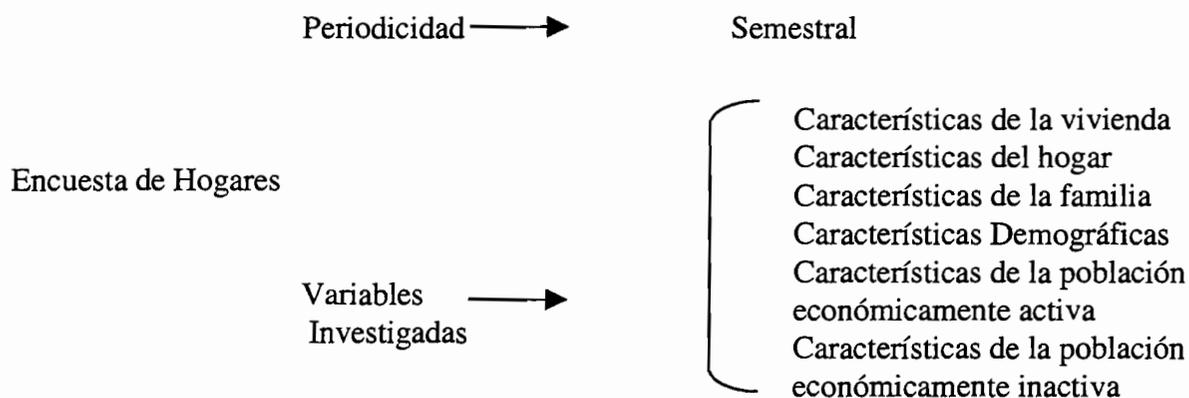
- ✓ **Encuestas Continuas:** Encuestas que manejan variables susceptibles de modificaciones significativas en corto plazo o variables coyunturales.
- ✓ **Encuestas no periódicas y complementarias:** Que incluyen temas de interés para el diseño de políticas económicas y sociales a ser estudiadas, periódicas o eventualmente.



Encuesta de hogares por muestreo

Es una encuesta periódica, que se realiza en el país de forma ininterrumpida desde el año 1967, y genera información sobre el tamaño de la Fuerza de Trabajo, proporcionan cifras relativas al desempleo y demás aspectos inherentes al mercado de trabajo venezolano, cubriendo las demandas de información, tanto a nivel nacional como internacional.

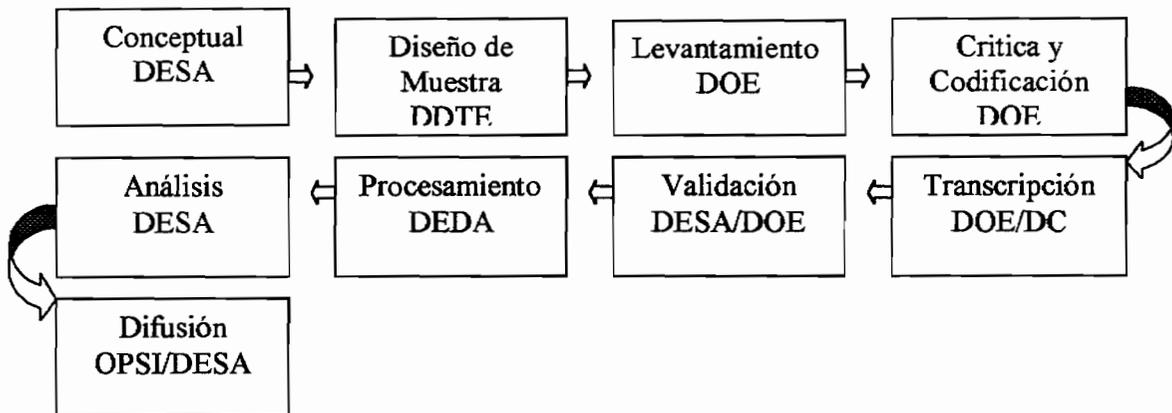
La Encuesta de Hogares por Muestreo constituye la principal fuente de datos que alimenta el Sistema de Encuestas a Hogares, dado la información que registra y la periodicidad con que se recoge y procesa la data.



En el diseño y ejecución de la Encuesta, participan diferentes direcciones, dentro de la estructura organizativa del INE, a saber:

- ✓ Dirección de Diseños y Técnicas Estadísticas (DDTE)
- ✓ Dirección de Estadísticas Sociales y Ambientales (DESA)
- ✓ Direcciones Estadales (DE)
- ✓ Dirección de Operaciones de Encuestas (DOE)
- ✓ Dirección de Computación (DC)
- ✓ Oficina de Promoción y Servicios de Información (OPSI)

A continuación se presenta un diagrama donde se contemplan las 8 grandes etapas que comprenden a la EHM y las direcciones que en cada una intervienen:



3. Aspectos Computacionales de la EHM.

Actualmente en el INE, los procesos computacionales que requiere la Encuesta de Hogares por Muestreo para la generación de resultados, no se encuentran integrados en un solo sistema, son aplicaciones independientes para cada uno de los procesos, y son llevados a cabo por direcciones diferentes dependiendo de la etapa que se ejecute.

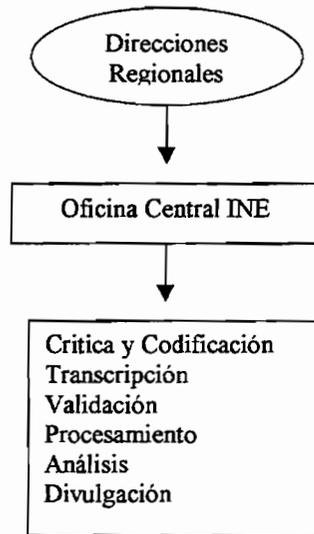
En tal sentido el INE emprendió un Proyecto de Actualización y Modernización de la Plataforma Informática, cuya filosofía se basa en el rediseño y migración de todas las aplicaciones estadísticas y administrativas a una plataforma web. La idea fundamental consistió en la desconcentración de los procesos a las diferentes regiones del país, (iniciándose con el proceso de transcripción) a fin de garantizar la oportunidad de la información generada por las diferentes encuestas.

Este proyecto no ha podido ser implementado por razones presupuestarias, debido a la gran inversión económica para la adquisición del Hardware y el desarrollo del Software.

3.1 Estrategia para la captura de datos

3.1.1 Centralización de los procesos.

Una vez diligenciados los cuestionarios en cada una de las direcciones regionales, en donde semanalmente se ejecuta el levantamiento de la encuesta de hogares, y aplicado los controles de calidad respectivos a los mismos, son enviados vía valija a la oficina central con sede en Caracas, para continuar con los siguientes procesos en la obtención de resultados de la encuesta, cabe resaltar que el único proceso descentralizado de la encuesta, es el levantamiento.



3.1.2 Crítica y codificación de la planillas

La crítica y codificación son dos procesos que se realizan simultáneamente, previos a la transcripción, y está orientado a examinar los datos contenidos en las planillas para detectar la mayor cantidad de inconsistencia, errores y omisiones que se puedan presentar, para luego corregirlas y depurar el dato. Este proceso permite afinar la información recabada en campo, sometiéndola a una revisión exhaustiva antes de que sea transcrita y procesada, lo cual, garantiza la minimización de errores en los procesos posteriores.

Los cuestionarios de la EHM están diseñados utilizando varios tipos de preguntas, respetando un orden de llenado. Así, tenemos preguntas abiertas, cerradas y de respuestas múltiples. La mayor parte del cuestionario corresponde a preguntas cerradas donde requieren respuestas estrictamente limitadas a las categorías establecidas en ellas. Estas categorías son identificadas a través de un código que debe ser trasladado al lugar correspondiente, lo que llamamos, preguntas precodificadas.

Las preguntas que requieren codificación -en estos momentos se realiza en forma manual- son las relacionadas con los grupos de ocupación y rama de actividad económica, así como el código de entidades federales y países.

N
ú
m
e
r
o

d
e

L
i
n
e
a

a. CARACTERÍSTICAS GENERALES							
<p style="text-align: center;">(18)</p> <p style="font-size: small;">Transcribe los códigos correspondiente al N° de línea, sexo y parentesco de todas las personas anotadas en el Cuestionario A, Sección III, Preguntas 7, 9 y 10</p> <table style="width: 100%; border: none;"> <tr> <td style="width: 50%; border: none;"> <p>Sexo</p> <p>1. Masculino</p> <p>2. Femenino</p> </td> <td style="width: 50%; border: none;"> <p>Parentesco</p> <p>1. Jefe o jefe(a) del hogar</p> <p>2. Esposa(o)</p> <p>3. Compañera(o)</p> <p>4. Hijos(as)</p> <p>5. Hijastros(as)</p> <p>6. Nietos(as)</p> <p>5. Yernos, nueros</p> <p>6. Padre, madre</p> <p>7. Suegro(a)</p> <p>8. Hermano(a)</p> <p>9. Cuñado(a)</p> <p>10. Sobrino(a)</p> <p>11. Tío</p> <p>12. Primo(a)</p> <p>13. Abuelo(a)</p> <p>14. Otro pariente</p> <p>15. No pariente</p> <p>15. Servicio domestico</p> <p>17. Familiares del servicio doméstico</p> </td> </tr> </table>	<p>Sexo</p> <p>1. Masculino</p> <p>2. Femenino</p>	<p>Parentesco</p> <p>1. Jefe o jefe(a) del hogar</p> <p>2. Esposa(o)</p> <p>3. Compañera(o)</p> <p>4. Hijos(as)</p> <p>5. Hijastros(as)</p> <p>6. Nietos(as)</p> <p>5. Yernos, nueros</p> <p>6. Padre, madre</p> <p>7. Suegro(a)</p> <p>8. Hermano(a)</p> <p>9. Cuñado(a)</p> <p>10. Sobrino(a)</p> <p>11. Tío</p> <p>12. Primo(a)</p> <p>13. Abuelo(a)</p> <p>14. Otro pariente</p> <p>15. No pariente</p> <p>15. Servicio domestico</p> <p>17. Familiares del servicio doméstico</p>	<p style="text-align: center;">(19)</p> <p style="font-size: small;">¿Cuál es la edad en años cumplidos de...?</p> <div style="border: 1px solid black; padding: 2px; font-size: small; text-align: center;"> <p>ANOTE "CERO" SI LA PERSONA ES MENOR DE UN AÑO</p> </div> <div style="border: 1px solid black; padding: 2px; font-size: small; text-align: center; margin-top: 10px;"> <p>Menores de 12 años</p> <p style="text-align: center;">↓</p> <p style="text-align: center;">(22)</p> </div>	<p style="text-align: center;">(20)</p> <div style="background-color: black; width: 100px; height: 30px; margin: 0 auto;"></div> <p style="font-size: small;">¿Cuál es la situación conyugal de...?</p> <p>1. Casado con cónyuge residente</p> <p>2. Casado con cónyuge no residente</p> <p>3. Unido con cónyuge residente</p> <p>4. Unido con cónyuge no residente</p> <p>5. Divorciado, separado</p> <p>6. Viudo</p> <p>7. Soltero</p>	<p style="text-align: center;">(21)</p> <p style="font-size: small;">Identifique los núcleos familiares en este hogar a partir de las personas registradas en el mismo; valide la conformación de los núcleos del hogar con el informante</p> <table style="width: 100%; border: none;"> <tr> <td style="width: 60%; border: none;"> <p style="font-size: small;">Cod. de identificación del núcleo familiar</p> <p style="font-size: small;">N°</p> </td> <td style="width: 40%; border: none;"> <p style="font-size: small;">1 1er. núcleo</p> <p style="font-size: small;">2 2do núcleo</p> <p style="font-size: small;">3 3er. núcleo</p> <p style="font-size: small;">4 4to. núcleo</p> <p style="font-size: small;">5 5to. núcleo</p> <p style="font-size: small;">0 Ningún núcleo</p> </td> </tr> </table> <p style="font-size: small; margin-top: 10px;">NEXO DEL NÚCLEO</p> <p style="font-size: small;">1. Padre, padastro, compañero, esposo</p> <p style="font-size: small;">2. Madre, madastra, compañera, esposa</p> <p style="font-size: small;">3. Hijo(a), hijastro(a)</p>	<p style="font-size: small;">Cod. de identificación del núcleo familiar</p> <p style="font-size: small;">N°</p>	<p style="font-size: small;">1 1er. núcleo</p> <p style="font-size: small;">2 2do núcleo</p> <p style="font-size: small;">3 3er. núcleo</p> <p style="font-size: small;">4 4to. núcleo</p> <p style="font-size: small;">5 5to. núcleo</p> <p style="font-size: small;">0 Ningún núcleo</p>
<p>Sexo</p> <p>1. Masculino</p> <p>2. Femenino</p>	<p>Parentesco</p> <p>1. Jefe o jefe(a) del hogar</p> <p>2. Esposa(o)</p> <p>3. Compañera(o)</p> <p>4. Hijos(as)</p> <p>5. Hijastros(as)</p> <p>6. Nietos(as)</p> <p>5. Yernos, nueros</p> <p>6. Padre, madre</p> <p>7. Suegro(a)</p> <p>8. Hermano(a)</p> <p>9. Cuñado(a)</p> <p>10. Sobrino(a)</p> <p>11. Tío</p> <p>12. Primo(a)</p> <p>13. Abuelo(a)</p> <p>14. Otro pariente</p> <p>15. No pariente</p> <p>15. Servicio domestico</p> <p>17. Familiares del servicio doméstico</p>						
<p style="font-size: small;">Cod. de identificación del núcleo familiar</p> <p style="font-size: small;">N°</p>	<p style="font-size: small;">1 1er. núcleo</p> <p style="font-size: small;">2 2do núcleo</p> <p style="font-size: small;">3 3er. núcleo</p> <p style="font-size: small;">4 4to. núcleo</p> <p style="font-size: small;">5 5to. núcleo</p> <p style="font-size: small;">0 Ningún núcleo</p>						

3.1.3 Captura de datos

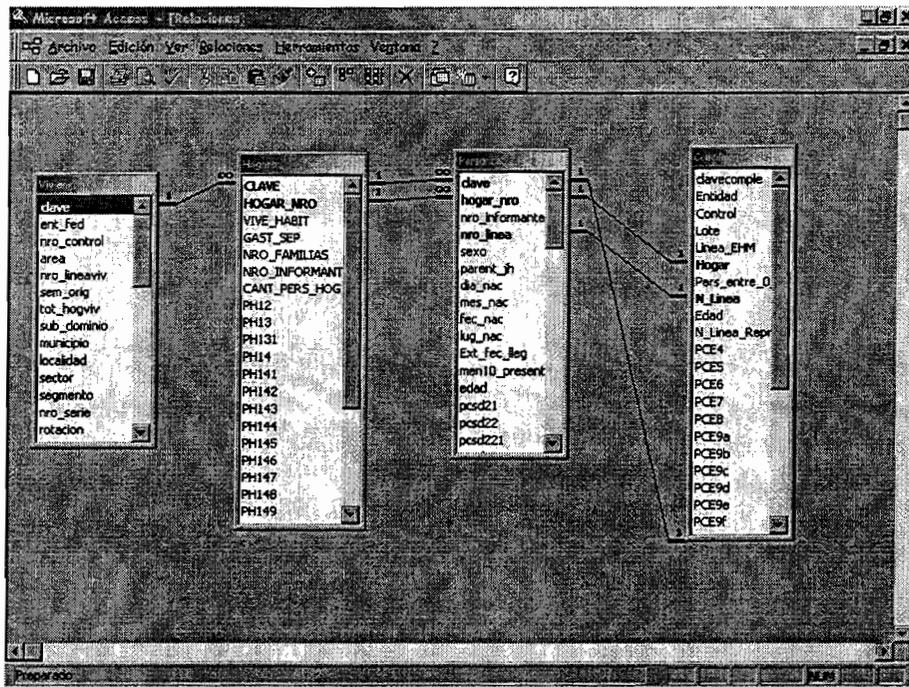
3.1.3.1 Diseño de estructura de datos

El Modelo lógico empleado en el sistema de captura de la EHM es de entidad-relación.

El sistema utiliza tres tablas de datos, una por cada unidad de análisis de la encuesta a saber:

- ✓ Tabla de Vivienda
- ✓ Tabla de Hogares
- ✓ Tabla de Personas

En el semestre en curso se levanta una encuesta complementaria de educación, cuya captura se hace en conjunto con la de la EHM, generándose una tabla adicional para dicha encuesta.



3.1.3.2 Desarrollo de la aplicación

El sistema de captura fue desarrollado en Visual Basic 6.0 con el manejador de base de datos Access 98 ó 2000.

El sistema de captura de la EHM cuenta con el siguiente menú:



- ✓ Encuesta: Opcion que permite transcribir las planillas de las EHM y las complementarias cuando existan.
- ✓ Conversión de los datos: Transforma los tablas de Access a formato dbf para su posterior manejo.
- ✓ Consolidación de la información : consolida las bases de datos en una sola y genera estructuras requeridas para el procesamiento de los datos

Pantalla de captura EHM

ENCUESTA DE HOGARES POR MUESTREO
CUESTIONARIO A

Opciones:

Sección I. Identificación

1. Entidad.....	16	4. N° de Línea (EH-III).....	100
2. N° de Control (UPM).....	1084	5. Semana del Levantamiento.....	22
3. Área (USM).....	07	6. Total de Hogares efectivos.....	1
7. Sub Distrito.....	0	11. Segmento.....	001
8. Municipio.....	01	12. N° de Serie (EH-III).....	01
9. Localidad.....	001	13. Rotación.....	1
10. Sector.....	000	14. Trimestre.....	1

Sección II. Control de la entrevista

1. Fecha de la Entrevista.....	01/07/2003	5. Apell. y Nomb. del Cod.....	125
3. Apell. y Nomb. del Sup.....	12	6. Informante.....	1
4. Apell. y Nomb. del Encuest.....	15		

Pantalla de captura, Encuesta Complementaria de Educación

3.1.3.3 Requerimiento de hardware y software.

Equipo mínimo para la instalación del sistema

- ✓ Pentium ó 586
- ✓ 64 mgbyte de Ram
- ✓ Windows 95 en adelante
- ✓ Espacio en disco de 100 mgbyte
- ✓ Acces 98 0 2000

Recursos existentes para el proceso de transcripción

- ✓ 7 Pc. Conectados en red
- ✓ 1 Pc como servidor
- ✓ 7 transcriptoras por turno
- ✓ Un supervisor de transcripción

3.1.3.4 Justificación para el desarrollo del sistema de captura

La EHM fue diseñada de modo tal que permite, incorporar encuestas complementarias en el momento que se estime conveniente para profundizar en temas de interés distintos a la fuerza de trabajo, pero vinculados a los hogares, de allí la necesidad de contar con un Sistema flexible que permita la inserción al módulo de transcripción de nuevas estructuras de captura de datos de esas encuestas.

3.2. Procedimientos para evaluar la consistencia de la información

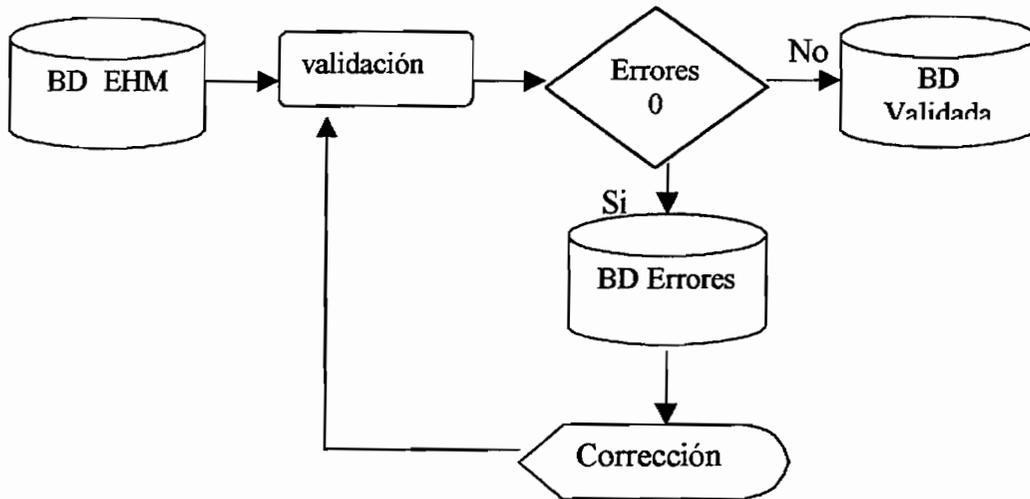
3.2.1 Validación

Se consideran dos niveles de validaciones:

- a. Validación en Línea: se realiza al momento de la entrada de datos y contempla:
 - ✓ Rango de datos: Valores permitidos para cada una de las variables
 - ✓ Consistencia y coherencia de los datos
 - ✓ Unicidad de registros
 - ✓ Normalización de registros
- b. Validación post-captura: Se realiza una vez consolidada la data, aplicándose el mismo plan de validación de la captura. Los errores encontrados son corregidos ejecutando un programa de corrección, repitiéndose el proceso hasta tanto quede depurada la base de datos. Esta validación se realiza por la limitación que tiene el sistema de captura de almacenar la información en una Base de Datos Unica durante el proceso de entrada. La información es almacenada en Bases de Datos locales y luego consolidada, mediante un proceso de respaldo y de carga de los datos.

El software para la validación y corrección de errores post-captura, esta desarrollado en FoxPro

El flujo de los datos una vez capturados es el siguiente:



3.3 Procesamiento de las bases de datos

3.3.1 Software utilizados y desarrollo de mismo.

Los programas para el procesamiento de datos han sido desarrollados en FoxPro. El desarrollo de los programas está a cargo de la Unidad de Consolidación y Consistencia de Resultados. Dado que la EHM es una investigación estadística, donde intervienen factores externos, susceptibles de ser modificados en el transcurso de la misma, se requiere un sistema flexible, para su procesamiento que permita introducir cambios en la ejecución de los programas empleados tantas veces sea necesario hasta tanto no se evalúen los resultados y se consideren estadísticamente satisfactorios.

3.3.2 Cálculo de los factores de expansión: Una vez depurados los datos se ejecuta este proceso, por el cual se obtiene la estimación poblacional a partir de los resultados de la muestra. Es un proceso completamente automatizado con un eficiente tiempo de respuesta, donde se aplican fórmulas estadísticas para los cálculos y se asignan pesos a los registros de las tablas de personas.

3.3.3 Generación de Tabulados: El plan de tabulación de la EHM está desarrollado en FoxPro, ya que este software permite el procesamiento de la data, de uno o varios de tabulados simultáneamente; ya sea en forma mensual, trimestral o semestral, así como por grupos de entidades y en tiempos muy cortos, lo que permite entregar la información para su análisis oportunamente, también permite la salida de los tabulados en diferentes formatos: Excel, Lotus, ASCII, dbf, lo que le da flexibilidad al usuario para hacer sus propios cálculos en el análisis

La Unidad de Consolidación y Consistencia de Resultados, brinda apoyo técnico al resto de los programas de la Dirección de Estadísticas, supliendo la necesidad de información de cada una de las encuestas del Sistema Encuestas a Hogares. Para tales demandas se procesa en SPSS, a través de ese sistema integrado se responde las solicitudes especiales de usuales externos.

ENCUESTA DE HOGARES (TABULADOS)

ENTIDAD: NACIONAL EXPANSION: EXPANDIDO

AÑO: FORMATO DE SALIDA: DBF

PERIODO: 1 SEMESTRE

ACEPTAR CANCELAR

O.C.E.I.

3.4. Imputación de datos

La Encuesta de Hogares por Muestreo, no realiza imputación de datos para la generación de los indicadores de la fuerza de trabajo.

Existen variables en la encuesta, donde no esta permitida la no respuesta, como los son las variables utilizadas para los cálculos de los factores de expansión (sexo y edad) y así como las preguntas que conforman el código sumario para la caracterización de las personas dentro o fuera de la Fuerza de Trabajo.

Existe para cada tabulado una categoría llamada "No declarado" donde se totaliza el número de personas que no respondieron las variables involucradas en el tabulado.

Los usuarios internos de las bases de datos de la EHM, como por ejemplo, el programa Indicadores de Condiciones de Vida desarrollan algoritmos para imputación de valores de la variable ingresos, se utiliza técnicas hot-deck y el sistema integrado STATA.

3.5. Integración de variables

Para el procesamiento de los tabulados de la EHM se incorporan a la base de datos variables generadas por algoritmos, como lo son: el peso de cada persona, situación de la personas en la fuerza de trabajo, informalidad, ingresos del hogar y otras más, quedando en ellas de forma permanente para uso y manipulación de los usuarios de las bases, anexándolas y describiéndoles en el diccionario de las bases de datos.

3.6. Integración de archivos de datos

Los archivos de la EHM que contienen la información de Vivienda, Hogares y Personas son generadas con una periodicidad semestral en formato dbf, por lo que pueden ser leídas y procesadas por los paquetes estadísticos, sin necesidad de ninguna transformación.

3.7. Políticas Institucionales de acceso a la base de datos

El INE, adelanta un proyecto de Políticas de Difusión de Información donde se prevé la comercialización electrónica de los datos. Mientras esto ocurre, los productos derivados de la encuesta se comercializan de manera directa, a través de la Oficina de Promoción y Atención al Usuario.

Existe un calendario preestablecido según la periodicidad, de difusión de la información, la cual puede ser: mensual, trimestral o semestral, estando disponibles los resultados mes y medio luego de culminado el mes de referencia.

Los productos comercializados hasta los momentos son:

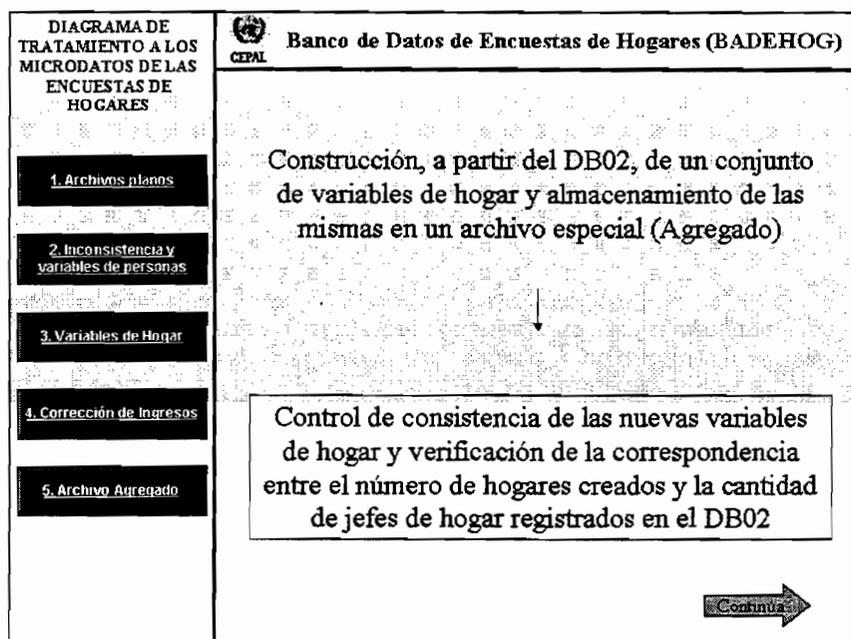
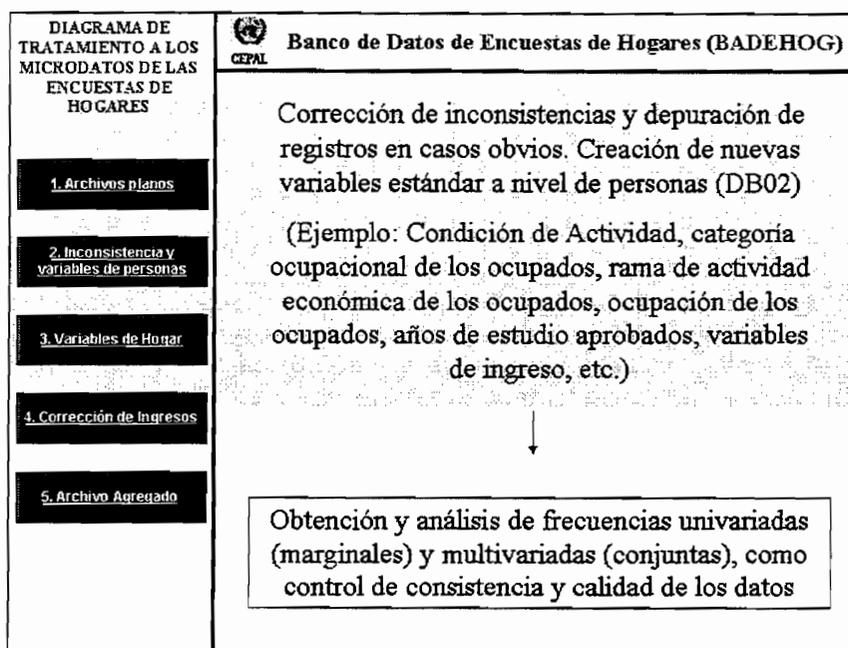
- ✓ Bases de datos Semestrales de la Encuesta de Hogares
- ✓ Libro 30 años de la Encuesta de Hogares por Muestreo
- ✓ CD ROOM 30 años de la Encuesta de Hogares por muestreo
- ✓ Libro Indicadores de la Fuerza de Trabajo Semestral
- ✓ Informes comparativos : Principales Indicadores de la Fuerza de Trabajo (mensual, trimestral y semestral)
- ✓ 108 Tabulados detallados de la Encuesta de Hogares por Muestro en formato impreso o electrónico.

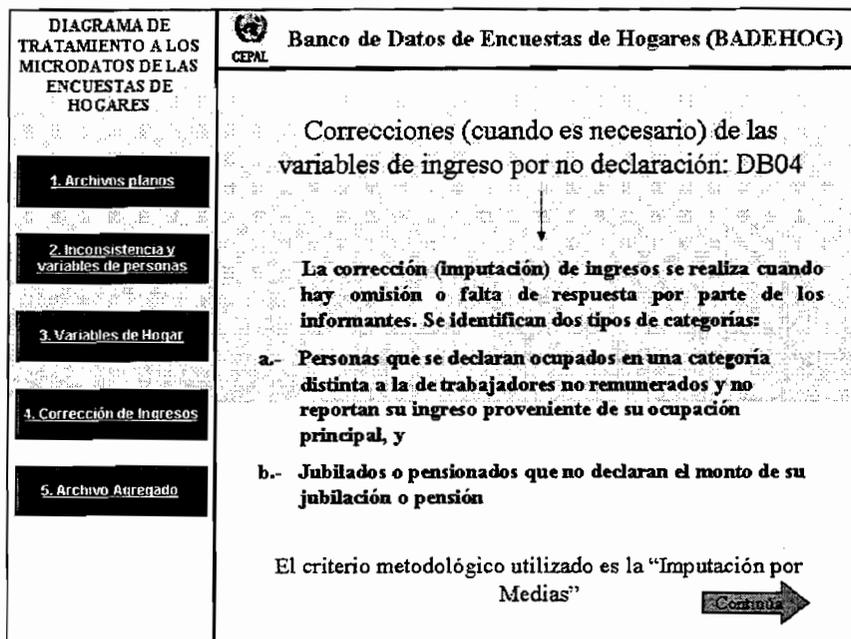
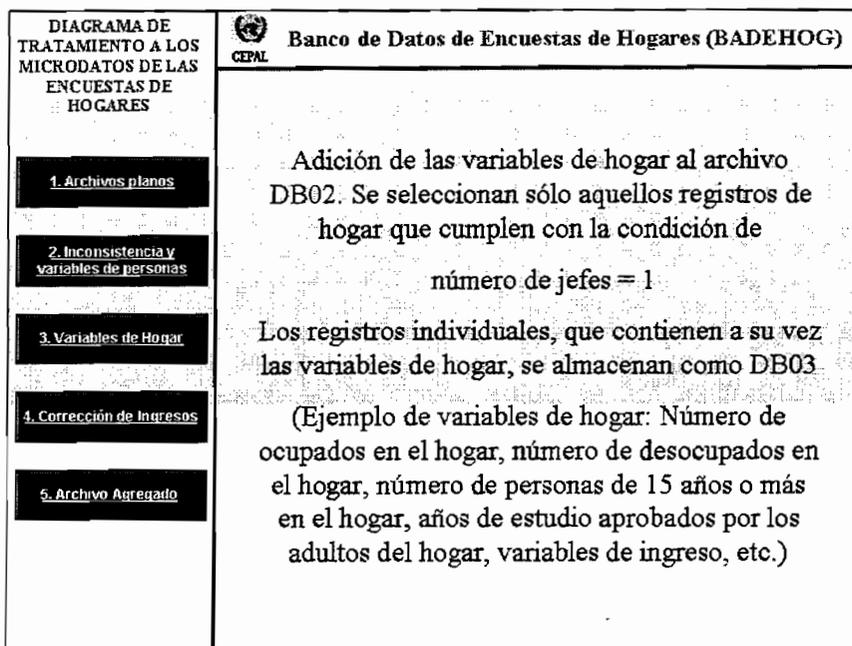
Actualmente contamos con la pagina web institucional, cuya dirección de acceso es: www.ine.gov.ve , donde se encuentran para consulta los Principales Indicadores de la Fuerza de Trabajo según la periodicidad disponible.

Sin embargo la información producida por la encuesta no se limita a la presentada en sus publicaciones, ya que el usuario puede solicitar procesamientos especiales para la obtención de información adicional.

**ESTANDARIZACIÓN DE LAS BASES DE DATOS DE
LAS ENCUESTAS DE HOGARES**

CEPAL





<p>DIAGRAMA DE TRATAMIENTO A LOS MICRODATOS DE LAS ENCUESTAS DE HOGARES</p> <p>1. Archivos planos</p> <p>2. Inconsistencia y variables de personas</p> <p>3. Variables de Hogar</p> <p>4. Corrección de Ingresos</p> <p>5. Archivo Agregado</p>	<p> Banco de Datos de Encuestas de Hogares (BADEHOG)</p> <ul style="list-style-type: none"> • Para los ocupados se imputa a cada persona el valor del ingreso medio reportado por aquellos de similares características, definidas en función del cruce simultáneo de cinco variables: jefatura del hogar, sexo, nivel educacional, categoría ocupacional y rama de actividad económica. • Para los jubilados o pensionados el procedimiento es similar al de los ocupados, sólo que las variables que intervienen son: jefatura del hogar, sexo y nivel educacional.
---	---

<p>DIAGRAMA DE TRATAMIENTO A LOS MICRODATOS DE LAS ENCUESTAS DE HOGARES</p> <p>1. Archivos planos</p> <p>2. Inconsistencia y variables de personas</p> <p>3. Variables de Hogar</p> <p>4. Corrección de Ingresos</p> <p>5. Archivo Agregado</p>	<p> Banco de Datos de Encuestas de Hogares (BADEHOG)</p> <p>Construcción de las variables de hogar creadas en la corrección y almacenamiento en un archivo especial (Agregado)</p> <p style="text-align: center;">↓</p> <p>Adición de las variables de hogar al archivo DB04.</p> <p>Los registros individuales, que contienen a su vez las variables de hogar, se almacenan como DB05</p> <p style="text-align: right;"></p>
---	--

<p>DIAGRAMA DE TRATAMIENTO A LOS MICRODATOS DE LAS ENCUESTAS DE HOGARES</p>	<p> Banco de Datos de Encuestas de Hogares (BADEHOG)</p>
<p>1. Archivos planos</p>	<p>Documentación de los archivos</p>
<p>2. Inconsistencia y variables de personas</p>	<ul style="list-style-type: none">• Manual del usuario (Ficha técnica de la encuesta, marco poblacional, diccionario de códigos de las variables)
<p>3. Variables de Hogar</p>	<ul style="list-style-type: none">• Metodología y cuestionario (Diseño muestral, definiciones y el cuestionario utilizado en el levantamiento de la encuesta)
<p>4. Corrección de Ingresos</p>	
<p>5. Archivo Asegado</p>	

BADEHOG

CEPAL



División de Estadística y Proyecciones Económicas
Unidad de Estadística Social

Forbes 2014
Período CPAL
English Version

BADEHOOG Banco de Datos de Encuestas de Hogares

CEPAL [Inicio](#) [Acceso de BADEHOOG](#) [Innovaciones](#) [Encuestas por país y año](#) [Sistema de notificación](#) [Ayuda](#)

Destacado

- AVANCE 2013**
Nuevas encuestas en curso
El primer de agosto, comenzó la recolección de datos en el primer trimestre de agosto correspondiente al año 2013 y 2014. Este periodo de encuestas se aplica a las encuestas nacionales.
- Encuestas 2013**
Badehog en inglés
Se implementa el sistema de actualización de datos en inglés en BADEHOOG.

Instrumento
Manual de Uso

Manual de uso
El manual de uso se encuentra en el menú de uso.

Ingrese al Sistema Generador de Frecuencias y Tablas haciendo click en un país.

- Argentina
- Bolivia
- Brazil
- Chile
- Colombia
- Costa Rica
- Ecuador
- El Salvador
- Guatemala
- Honduras
- Brasil
- Nicaragua
- Panamá
- Paraguay
- Perú
- República Dominicana
- Uruguay
- Venezuela

División de Estadística y Proyecciones Económicas - Unidad de Estadística Social

BADEHOG Banco de Datos de Encuestas de Hogares

Acerca de BADEHOG

Este Banco de datos, implementado por la Unidad de Estadística Social de la División de Estadística y Proyecciones Económicas, contiene información de las encuestas de hogares inventariadas en los países de América Latina desde alrededor de 1980 en adelante. Para la mayoría de las encuestas, proporciona antecedentes relativos a la metodología y manuales de usuario, como también los recuadros correspondientes.

Para la obtención de resultados de las encuestas se dispone de dos módulos, a saber:

1. Un Sistema Generador de Frecuencias y Tablas, que genera tablas en línea y en tiempo real.
2. Un módulo para Solicitud de Variables, que ofrece al usuario la opción de seleccionar y recibir bajo ciertos restricciones, indicadores que requieren procesamiento de acuerdo a su conveniencia.

Para el gobierno de BADEHOG se ha contado con la acostumbrada e invaluable colaboración de las Direcciones Nacionales de Estadística y demás instituciones de los países de la región encargadas del diseño y ejecución de las encuestas de hogares de países cubiertos.

Debido a restricciones restrictivas de algunos gobiernos al uso de la información disponible en el BADEHOG, todas las encuestas que conforman este banco de datos son en uso exclusivo del personal de la CEPAL, y bajo ninguna circunstancia pueden ser compartidas con personas ajenas a la Comisión sin antes recibir autorización expresa de su administrador: badehog@cepal.org

Este sitio se encuentra en permanente actualización. Antes de hacer uso de la información residente en el Banco, siempre revise las actualizaciones.

División de Estadística y Proyecciones Económicas - Unidad de Estadística Social

Encuestas de hogares por país y por año

	Argentina	Bolivia	Brasil	Chile	Colombia	Costa Rica	Ecuador	El Salvador	Guatemala	Honduras	México	Nicaragua	Panamá	Paraguay	Perú	República Dominicana	Uruguay	Venezuela
Argentina	x																	
Bolivia		x																
Brasil			x															
Chile				x														
Colombia					x													
Costa Rica						x												
Ecuador							x											
El Salvador								x										
Guatemala									x									
Honduras										x								
México											x							
Nicaragua												x						
Panamá													x					
Paraguay														x				
Perú															x			
República Dominicana																x		
Uruguay																	x	
Venezuela																		x

División de Estadística y Proyecciones Económicas - Unidad de Estadística Social

Chile: Encuesta CASEN
Noviembre 1998
 (Muestra representativa por comunas)
 Personas
 De 1992 a 1998

CASEN	Frecuencia	Porcentaje	Cumulativa Frecuencia	Cumulativa Porcentaje
1	229374	49.46	229374	49.46
2	128599	27.90	357973	77.36
3	688849	14.81	1046822	92.17
4	267881	58.07	1314703	99.97
5	6219	0.13	1320922	99.99
6	23064	0.50	1343986	100.00
7	7668	0.17	1351654	100.00
8	3513	0.08	1355167	100.00
9	1319	0.03	1356486	100.00

Universe:
 EDAD mayor que 15
 This request took 4.08 seconds of real time (at 2 build 1998).



 Portal de Datos: [Portal CPAL](#) [English Version](#)

BADEHOG Banco de Datos de Encuestas de Hogares

[Inicio](#) [Acerca de BADEHOG](#) [Innovaciones](#) [Encuestas por país y año](#) [Recopilación de variables](#)

[Ayuda](#) [Contacto](#) [Sugerencias](#)

Chile
Variables presenciales
Formas

Directa **Encuesta CASEN**
(En ingresos agregados)

Año: 1998
 Población: Residentes
 Nivel: Españoles
 Modalidad: [Habitación y Colectividad](#)
[Manual del Usuario](#)

YANZ Dividendo por acciones
YANZAL Dividendo por acciones
YANZGSH Alquiler pagurado
YANZM Alquiler pagurado
YANZEL Alquiler pagurado
YANT Ingreso autónomo
YANTAL Ingreso autónomo
YANTGSH Ingreso autónomo

Según por orden Alfabético Según manual

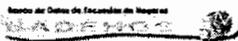
Porcentaje: S No
 Índice: S No
 Origen: S No
 Salto: Usar Normal
 Ent:

Procesar

Región: Variable: Subíndice: Columna: Orden: Ordenamiento: Ordenamiento:

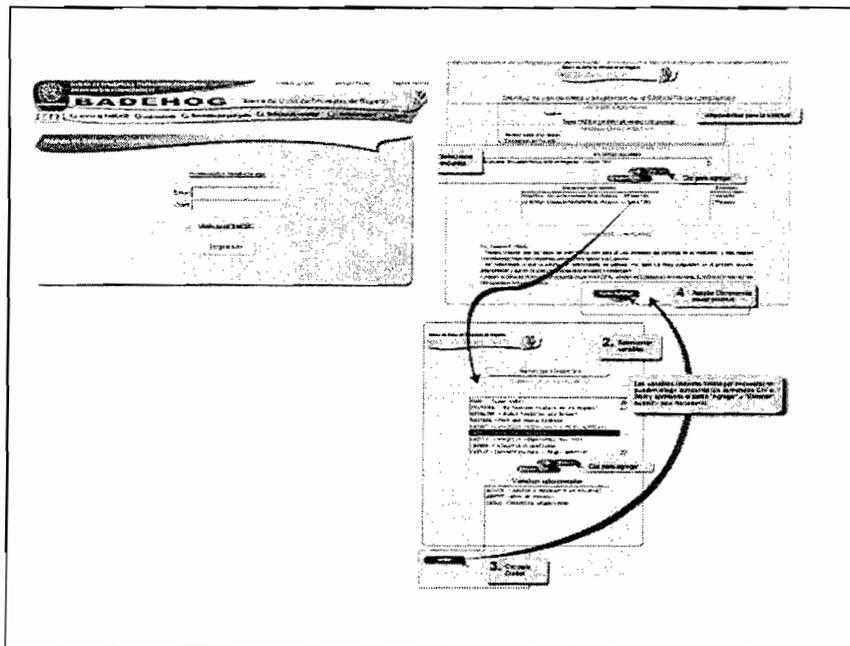
Ordenamiento: Ordenamiento: Ordenamiento: Ordenamiento: Ordenamiento: Ordenamiento:

División de Estadística y Proyecciones Económicas - Unidad de Estadística Social



Chile: Encuesta CASEN
Noviembre 1998
(Resaltados y sumatorias por columna)
 Porcentaje

INDICADOR	CASEN										
	1	2	3	4	5	6	7	8	9	10	Total
	YANT	YANTL	YANTGSH	YANTM	YANTEL	YANTGSH	YANTM	YANTL	YANTGSH	YANTM	YANTL
	Presencial	Presencial	Presencial	Presencial	Presencial	Presencial	Presencial	Autónomo	Presencial	Presencial	
1	26,927.30		27,891.80	44,489.02		41,692.61					38,114.45
2	42,712.01	81,362.11	50,932.34	103,093.25		71,054.94	10,000.00	96,779.36	34,000.00		81,110.31
3	124,218.79	489,471.14	216,159.25	297,875.27	135,467.17	104,821.74	101,154.79	279,405.84	2,650,000.00		264,309.49
Total	116,815.89	886,032.00	205,739.18	287,481.30	135,607.17	94,484.30	99,768.15	273,031.20	2,415,954.89		249,417.31



**CENSUS AND SURVEY PROCESSING SYSTEM
(CSPRO)**

SERPRO S.A.

ÍNDICE

	<u>Página</u>
CSProm.....	303
Diseñador Integrado de Aplicaciones.....	303
Definición de Diccionarios.....	303
Aplicaciones Inteligentes de Entradas de Datos.....	303
Consistencia e Imputación Batch	304
Batch.....	304
Tabulaciones.....	304
Utilitarios.....	304
CAPI.....	305
Sistema de Entrada Remoto	305

CSPro

- Diseñado para ser usado en Censos Y Encuestas
- Fácil de usar
- Implementado en plataforma Windows
- Diseño modular
- Definición única de archivos para todos los módulos
- Lenguaje común para todos los módulos
- Aplicaciones de ISSA/IMPS fácilmente transportables
- Más de 7.000 usuarios de más de 170 países
- Respaldo por el U.S. Census Bureau

Diseñador Integrado de Aplicaciones

- Diseñador gráfico *amistoso* para aplicaciones
- Diseño de Diccionarios, Forms (pantallas de ED), Tablas y Lógica con un solo diseñador y en forma simultánea
- Desarrollo de múltiples aplicaciones simultáneas
- Forms son generados automáticamente con base en diccionario o personalizados
- Técnica de “*drag & drop*” para diseño de forms de ED y tablas
- Lógica de la aplicación es escrita en lenguaje CSPro usando un editor integrado con compilación en línea
- Lógica asociada a cuestionario, forms, grupos y campos
- Cuatro eventos en el tiempo de ejecución de los procedimientos (relativos a cada objeto/campo): PreProc, OnFocus, KillFocus, PostProc

Definición de Diccionarios

- Un Diccionario (DD) describe la organización o estructura física de un caso o cuestionario
- Define cada uno de los tipos de registros de que se compondrá un caso
- Define cada una de las variables (ítems) que cada registro tiene
- Define los rangos o categorías de cada variable
- Constituye una muy buena documentación de los archivos de datos

Aplicaciones Inteligentes de Entrada de Datos

- Ingresa, Recupera o Modifica casos con la misma aplicación
- Verifica datos con comparación y corrección inmediata de errores
- Ilimitado número de forms de entrada de datos
- Scroll automático –horizontal y vertical- de los forms, sin limitaciones de tamaño
- Los forms pueden contener campos de distintos registros

- Los forms pueden contener matrices de dos dimensiones para ingreso de registros de ocurrencia múltiple (grilla)
- Aplicación contiene chequeos de consistencia y esquema de saltos automáticos tan complejos como se quiera
- Lógica es distribuida en cuatro eventos relativos al ingreso de cada campo: PreProc, OnFocus, KillFocus y PostProc
- Mensajes de error diseñados por el usuario
- Uso de Archivos o tablas externas a la aplicación en chequeos de consistencia
- Aplicación de entrada de datos puede ser ejecutada en batch para generación de listado de errores finales
- Producción automática de estadística de digitadores

Consistencia e Imputación Batch

- Poderoso lenguaje estructurado
- Chequeo de variables a nivel de registro y entre registros

Batch

- Chequeo de la estructura de casos
- Imputación de variables modificando valores originales o creando nuevas variables
- Generación de informes automáticos o diseñados ad-hoc
- Acceso a múltiples archivos (lectura/escritura) secundarios para apoyo en proceso de edición

Tabulaciones

- Tabulación de complejas estructuras de datos sin necesidad de programar
- Producción de sofisticadas tablas sin necesidad de programar
- Poderosa interfaz gráfica para desarrollo de tablas, con técnica de *drag & drop* directamente desde el diccionario
- Los formatos de las tablas son dibujadas en la pantalla tal y cual serán impresas con posterioridad
- Selección del universo de la tabla mediante expresiones lógicas formuladas por medio de la IGU
- Definición de unidad a tabularse, con o sin peso (factor de expansión de la muestra)
- Generación de tablas a niveles de áreas geográficas y posterior cálculo de agregados a distintos niveles
- Cálculo de variadas estadísticas
- Cálculo de errores muestrales
- IGU orientada a eliminar el requerimiento de un programador en la obtención de tablas

Utilitarios

- Distribución de frecuencias de variables
- Ordenamiento de casos por múltiples llaves

- Exportación de datos y generación de diccionarios (*codebooks*) a SPSS, SAS y STATA
- Reformateo de archivos de datos
- Comparación de archivos de datos

CAPI

- Notorio mejoramiento en calidad de la información
 - Generación dinámica de la pregunta usando información previamente recogida permite precisar y personalizar texto
 - Espacio ilimitado para desplegar texto permite mejorar la formulación de las preguntas
 - Comunicación fina con entrevistador/a permite instruir y/o recordar en momento oportuno determinados tópicos
 - Formulación de las preguntas es uniforme, eliminándose en gran medida cierto sesgo personal introducido por algunos entrevistadores
 - El idioma en que se formula la pregunta puede ser cambiado en cualquier momento durante la entrevista
 - Cada respuesta del entrevistado/a puede ser confrontada con cualquier información relevante ya ingresada, detectando errores de consistencia cuando es fácil corregir
 - La consistencia en línea permite también corregir errores originados por entrevistador/a y no de imputar posteriormente
 - El flujo de la entrevista es manejado enteramente por el computador basado en las respuestas de los entrevistados; ello implica que los *valores faltantes* –*Missing Values*– son prácticamente eliminados de la encuesta
 - En encuestas complejas y largas, la duración de la entrevista es disminuida en forma importante, y por ende también el hastío y cansancio del entrevistado
- Ahorros inherentes a la operación CAPI
 - 100 % costo impresión de los cuestionarios
 - Costo del diseño físico de los cuestionarios (el diseño lógico siempre es necesario)
 - Costo de instrucción de los digitadores
 - Costo de digitación y corrección/imputación de datos
 - Reducción del tiempo de la entrevista por caso
 - Reducción del tiempo requerido para la obtención de la base de datos final, lista para su análisis

Sistema de Entrada Remoto

- Basado en CPro mono-usuario
- Usa esquema Cliente-Servidor conectados a través de una intranet o internet
- Características principales del servidor son:
 - Corre actualmente en una plataforma Windows y próximamente se implementará bajo Unix/Linux

- Múltiples servidores en uno o más computadores para un mismo proyecto
- Conexión a cualquier base de datos relacional a través de conector (driver) ODBC
- Implementación de último algoritmo criptográfico simétrico (Advanced Encryption Standard AES 128 o 256 bits)
- Características principales del cliente son:
 - La misma aplicación de mono-usuario puede correr como remota
 - Refresco automático de las aplicaciones al momento de conectarse (asegura permanente consistencia de las versiones de todos los clientes)
 - Acceso restringido a casos de otros clientes
 - Los datos son transmitidos al final de cada caso u opcionalmente, al final de la sesión
 - Las aplicaciones son mantenidas en formato ZIP y encriptadas, por lo cual están protegidas de cualquier modificación

CSPRO fue desarrollado por la Oficina del Censo de los Estados Unidos, ORC Macro y SERPRO S.A. y representa la integración de los dos sistemas, ISSA e IMPS.

PROGRAMA PARA EL MEJORAMIENTO DE LAS ENCUESTAS Y LA MEDICION DE LAS
CONDICIONES DE VIDA EN AMERICA LATINA Y EL CARIBE (MECOVI)
BID – BANCO MUNDIAL - CEPAL

11° TALLER REGIONAL

Metodologías para la Integración de Bases de Datos de Encuestas de Hogares

Lima, Perú, 16 al 18 de julio de 2003

INEI

CEPAL

Lista de Participantes

ARGENTINA

Elsa Beatriz López
Responsable líneas de desarrollo y
Programación de la Encuesta Permanente de Hogares
Instituto Nacional de Estadística y Censos
Julio A. Roca 609
Buenos Aires – Argentina
Tel.: 54-11 4349 9206
E-mail: blope@indec.mecon.gov.ar

Noemí Raquel Salva
Analista Programador de la Dirección de Estudios
e Ingresos
Instituto Nacional de Estadística y Censos
Av. Julio A. Roca 609
Buenos Aires – Argentina
Tel.: 54-11 43499240
Fax: 54 -11 43499886
E-mail: rsalv@indec.mecon.gov.ar

BOLIVIA

Freddy O. Ochoa Cabrera
Responsable Informática MECOVI
Instituto Nacional de Estadística INE
Calle Carrasco 1391
La Paz – Bolivia
Tel.: 5912 2222333
Fax: 5912 2222885
E-mail: fochoa@ine.gov.bo

CHILE

Héctor Osvaldo Morales H.
Jefe de Departamento de Atención al Usuario y
Difusión
Instituto Nacional de Estadísticas

Av. Bulnes 418
Santiago – Chile
Tel.: 562 3667407
E-mail: hector.morales@ine.cl

Inés Cáceres R.
Jefe Subdepto. Encuestas de Hogares
Instituto Nacional de Estadísticas
Av. Bulnes 418
Santiago- Chile
Tel.: 562 3667606
Fax: 562 6722460
E-mail: ines.caceres@ine.cl

COLOMBIA

Edgar Alexander Prieto M.
Jefe Oficina de Sistemas
Departamento Administrativo Nacional
De Estadística – DANE
Transversal 45 # 26 –70 Int. 1 CAN
Bogotá D.C. – Colombia
Tel.: 57 1 5978350
Fax: 57 1 5978306
E-mail: eaprietom@dane.gov.co

COSTA RICA

Carlos Loría Loría
Técnico en Informática
Instituto Nacional de Estadística y Censos (INEC)
De la Rotonda de la Bandera 450 mts. Oeste
Sobre calle Los Negritos, Edificio Ana Lorena
Tel.: 506 224 3300
Fax: 506 281 2263
E-mail: carlos.loria@inec.go.cr

Randall Céspedes Villalobos
DBA, Profesional Licenciado en Informática
Instituto Nacional de Estadística y Censos (INEC)

De la Rotonda de la Bandera 450 mts. Oeste
Sobre calle Los Negritos
Edificio Ana Lorena , Mercedes de Montes de Oca
San José – Costa Rica
Tel.: (506) 224-3000
Fax: (506) 281-2263
E-mail: randall.cespedes@inec.go.cr

ECUADOR

Carlos Iván Obando R.
Especialista en Sistemas Administrativos
Instituto Nacional de Estadística y Censos INEC
Juan Larrea No. 15-36 y José Riofrío
Quito – Ecuador
Tel.: 593 2 2544326
Fax: 593 2 2528097
E-mail: ivanobando@hotmail.com

GUATEMALA

Vivian María Guzmán Q.
Analista de Resultados
Instituto Nacional de Estadística
8 calle 9-55 zona 1
Guatemala – Guatemala
Tel.: 502 2320962
Fax: 502 2320970
E-mail: guzmanvivian@yahoo.com

HONDURAS

Ivonne Ramírez
Gerente
Instituto Nacional de Estadística (INE)
Boulevard Suyapa, Col. Florencia Sur
Tegucigalpa – Honduras
Tel.: 504 2397161
Fax: 504 2398612
E-mail: iramirez@ine-hn.org

Dennis Alexander Martínez M.
Técnico en Procesamiento de Datos
Instituto Nacional de Estadística (INE)
Boulevard Suyapa, Col. Florencia Sur
Tegucigalpa – Honduras
Tel.: 504 23971 / 23954
Fax: 504 2398612
E-mail: dmartinezm@unitec.edu

MÉXICO

Ricardo Luján Salazar
Coordinador de desarrollo de Sistemas de Base de Datos

Instituto Nacional de Estadística, Geografía e Informática (INEGI)
Héroe de Nacoziari Sur 2301, Fracc. Jardines del Parque
Aguascalientes – México
Tel. : (52-449) 4424796
Fax: (52-449) 55987738
E-mail: ricardo.lujan@inegi.gob.mx

Benito Durán Romo
Subjefe de Departamento
Instituto Nacional de Estadística, Geografía e Informática (INEGI)
Héroe de Nacoziari Sur 2301, Fracc. Jardines del Parque
Aguascalientes – México
Tel.: 52 449 9787912
Fax: 52 449 9787912
E-mail: bduran@dge.inegi.gob.mx

NICARAGUA

Santiago Mejía Solís
Responsable Depto. Procesamientos
Instituto Nacional de Estadísticas y Censos INEC
Residencia Los Arcos, Frente a Hospital
Antonio Lenín Fonseca
Managua – Nicaragua
Tel.: 505 2663955
Fax: 505 2663021
E-mail: smejiasolis@inec.gob.ni

PANAMA

Marcelino Peñalba García
Analista - Programador
Contraloría General de la República
Estadística y Censo
Panamá
Tel.: 507 2104823
Fax: 507 2104821
E-mail: marcelinop@contraloria.gob.pa

PARAGUAY

Oscar Esteban Correa Paredes
Analista de Sistemas y Programador
Dirección General de Estadística, Encuestas y Censos
Naciones Unidas y Saavedra
Asunción – Paraguay
Tel.: 595-21 551016 / 670935 / 205424
Fax: 595 21 508493
E-mail: oco@dgeec.gov.py

PERU

Nancy Hidalgo Calle
Directora Ejecutiva de Censos y Encuestas de Hogares
Instituto Nacional de Estadística e Informática
Gral. Garzón No. 658, Jesús María
Lima 11 – Perú
Tel.: 51 1 4333118
Fax: 51 1 4333118
E-mail: nhidalgo@inei.gob.pe

Pamela Cueto Paz Vergara
Directora Técnica de Difusión
Instituto Nacional de Estadística e Informática
Gral. Garzón No. 658, Jesús María
Lima 11 - Perú
Tel.: 51 1 4336565
Fax: 51 1 4311340
E-mail: pamela.cueto@inei.gob.pe

Juan Roberto Bustamante Mejía
Director Técnico de Informática
Instituto Nacional de Estadística e Informática
Gral. Garzón No. 658, Jesús María
Lima 11 - Perú
Tel.: 51 1 4338327
Fax: 51 1 4333159
E-mail: jbustamante@inei.gob.pe

Pedro Llontop Ledesma
Subgerente Encuestas
Instituto Cuánto
Plaza del Ovalo 203-B
San Isidro, Lima – Perú
Tel: 511 4425460
Fax: 511 4423421
E-mail: pllontop@cuanto.org

Juan José Manco Y.
Investigador Analista
Instituto Cuánto
Perricholi 109
San Isidro, Lima – Perú
Tel: 511 4425460
Fax: 511 4425460
E-mail: jjmanco@cuanto.org

REPUBLICA DOMINICANA

Francisco Cáceres Ureña
Encargado del Depto. de Encuestas
Oficina Nacional de Estadística

Edificio de Oficinas Gubernamentales Juan Pablo Duarte
Av. Francia Esquina Leopoldo Navarro
Santo Domingo – República Dominicana
Tel.: 1 809 6827777
Fax: 1 809 6868276
E-mail: fcaceres@profamilia.org.do

Dilia Margarita Cruz
Subdirectora de Encuestas Económicas
Banco Central de la República Dominicana
Av. Pedro Henriquez Ureña esquina Leopoldo Navarro
Santo Domingo – República Dominicana
Tel.: 1 809 6850783
Fax: 1 809 6827666
E-mail: d.cruz@bancocentral.gov.do

URUGUAY

Carlos Eduardo Calvo Peluffo
Subdirector Div. Estadísticas Sociodemográficas
Instituto Nacional de Estadística
Río Negro 1520
Montevideo – Uruguay
Tel.: 5982 9008902
Fax: 5982 9014653
E-mail: ccalvo@ine.gub.uy

VENEZUELA

Irina de Jesús de la Rosa S.
Planificador II
Instituto Nacional de Estadística
Av. Boyacá Edif. Fundación La Salle piso 1
Maripérez – Caracas – Venezuela
Tel.: 58 212 7930580
Fax: 58 212 7930580
E-mail: irinadelarosa@hotmail.com

CEPAL

Juan Carlos Feres
Coordinador MECOVI – CEPAL
División de Estadística y Proyecciones Económicas
Av. Dag Hammarskjöld 3477
Vitacura, Santiago – Chile
Tel.: 562 210 2408
Fax: 562 210 2472
E-mail: jferes@eclac.cl

Fernando Medina
Asesor Regional
División de Estadística y Proyecciones Económicas
CEPAL
Av. Dag Hammarskjöld 3477

Vitacura, Santiago – Chile
 Tel.: 562 210 2417
 Fax: 562 210 2472
 E-mail: fmedina@eclac.cl

María de la Luz Avendaño
 Coordinadora Banco de Datos de Encuestas de Hogares
 División de Estadística y Proyecciones Económicas
 CEPAL
 Av. Dag Hammarskjöld 3477
 Vitacura, Santiago – Chile
 Tel.: 562 210 2441
 Fax: 562 210 2472
 E-mail: mavendano@eclac.cl

Carlos Howes
 Experto en Procesamiento Electrónico de Datos Estadísticos
 División de Estadística y Proyecciones Económicas
 CEPAL
 Av. Dag Hammarskjöld 3477
 Vitacura, Santiago – Chile
 Tel.: 562 2102178
 Fax: 562 2102472
 E-mail: chowes@eclac.cl

BID

José Antonio Mejía Guerra
 Coordinador MECOVI – BID
 1300 New York Avenue, N.W. Washington D.C.
 20577, USA
 Tel.: 1 202 6233713
 Fax: 1 202 6233299
 E-mail: joseam@iadb.org

SERPRO S.A.

Julio Ortúzar
 Santiago, Chile
 Tel.: 562 3358253
 E-mail: joc@serpro.com

Rubén Hume
 Gerente Técnico
 Santiago, Chile

SPSS

Roy Molina
 Telemática (SPSS)
 Lima, Perú

OBSERVADORES INEI – PERU

Rosario Céspedes Klescovich
 Directora Nacional de Censo y Encuestas

Elcira Azañero Terrones
 Directora Ejecutiva de Desarrollo de Sistemas

Isabel Valladares Alcántara
 Profesional de la Dirección Nacional de Censos y Encuestas

Agustín Beretta Enciso
 Administrativo de Servicios Informáticos

Martha Tizón Palacios
 Metodóloga de la Dirección Nacional de Censos y Encuestas

Luis Ulloa Jesús
 Profesional Dirección Técnica de Demografía e Indicadores Sociales

Prudencia Javier Rimey
 Profesional de la Dirección Nacional de Censos y Encuestas

Lucía Gaslac Torres
 Metodóloga de la Dirección Nacional de Censos y Encuestas

