



Universidad de San Carlos de Guatemala
Facultad de Ingeniería
Escuela de Ingeniería en Ciencias y Sistemas

**TECNICAS DE PROCESAMIENTO DE DATOS
EN CENSOS Y ENCUESTAS**

ALEJANDRO MARDOQUEO PONCE RODAS

ASESORADO POR ING. CRESENCIO GERTRUDIS CHAN CANEK

GUATEMALA, FEBRERO DE 2005

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA



FACULTAD DE INGENIERÍA

**TÉCNICAS DE PROCESAMIENTO DE DATOS
EN CENSOS Y ENCUESTAS**

TRABAJO DE GRADUACIÓN

PRESENTADO A JUNTA DIRECTIVA DE LA
FACULTAD DE INGENIERÍA
POR

ALEJANDRO MARDOQUEO PONCE RODAS

ASESORADO POR ING. CRESENCIO GERTRUDIS CHAN CANEK

AL CONFERÍRSELE EL TÍTULO DE
INGENIERO EN CIENCIAS Y SISTEMAS

GUATEMALA, FEBRERO DE 2005

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA



FACULTAD DE INGENIERÍA

NÓMINA DE JUNTA DIRECTIVA

DECANO	Ing. Sydney Alexander Samuels Milson
VOCAL I	Ing. Murphy Olympo Paiz Recinos
VOCAL II	Lic. Amahán Sánchez Álvarez
VOCAL III	Ing. Julio David Galicia Celada
VOCAL IV	Br. Kenneth Issur Estrada Ruiz
VOCAL V	Br. Elisa Yazminda Vides Leiva
SECRETARIO	Ing. Carlos Humberto Pérez Rodríguez

TRIBUNAL QUE PRACTICÓ EL EXAMEN GENERAL PRIVADO

DECANO	Ing. Sydney Alexander Samuels Milson
EXAMINADOR	Ing. Luis Alberto Vettorazzi España
EXAMINADOR	Ing. Marlon Pérez Turk
EXAMINADOR	Ing. Edgar René Ornelyz
SECRETARIO	Ing. Pedro Antonio Aguilar Polanco

HONORABLE TRIBUNAL EXAMINADOR

Cumpliendo con los preceptos que establece la ley de la Universidad de San Carlos de Guatemala, presento a su consideración mi trabajo de graduación titulado:

TÉCNICAS DE PROCESAMIENTO DE DATOS EN CENSOS Y ENCUESTAS

Tema que me fuera asignado por la Escuela de Ingeniería en Ciencias y Sistemas de la Facultad de Ingeniería con fecha febrero de 2004.

Alejandro Mardoqueo Ponce Rodas

ÍNDICE

ÍNDICE DE ILUSTRACIONES.....	IV
GLOSARIO.....	VI
RESUMEN.....	VIII
OBJETIVOS.....	IX
INTRODUCCIÓN.....	XI

1 IMPORTANCIA Y PROPÓSITO DE LOS CENSOS Y ENCUESTAS

1.1 Sistema estadístico integrado.....	1
1.1.1 Objetivos de un sistema integrado.....	2
1.1.2 Recolección y publicación de datos.....	3
1.1.3 Elementos de un sistema estadístico integrado.....	8
1.1.4 Factores para desarrollar un sistema integrado.....	8
1.1.5 Responsabilidades de un sistema integrado.....	12
1.1.6 Integración de censos y encuestas por muestreo.....	14
1.2 Papel y objetivos de los censos y encuestas.....	15
1.2.1 Papel para responder a requerimientos estadísticos	15
1.2.2 Objetivos estadísticos.....	17
1.2.3 Objetivos operativos.....	20
1.2.4 Consideraciones al ejecutar un censo o encuesta...	22
1.2.5 La precisión, un objetivo del censo y de las encuestas.....	26
1.2.6 Importancia de actualizar los tópicos.....	28
1.3 Etapas preliminares al procesamiento.....	29
1.3.1 Planificación y presupuestos.....	29
1.3.2 Definición de temática por investigar.....	30
1.3.3 Actualización cartográfica.....	30
1.3.4 Cargas de trabajo.....	31

1.3.5 Reproducción de material.....	31
1.3.6 Empadronamiento.....	31
2 TEORÍA DEL PROCESAMIENTO DE DATOS	
2.1 Procesos implicados en el procesamiento de datos.....	33
2.1.1 Consideraciones básicas.....	33
2.1.2 Crítica.....	37
2.1.3 Codificación.....	41
2.1.4 Métodos de conversión de datos	42
2.1.5 Programas de computadora.....	45
2.2 Administración del procesamiento.....	48
2.2.1 Pruebas de sistema.....	48
2.2.2 Requisitos y funciones del personal.....	51
2.2.3 Sistemas de control.....	52
2.2.4 Ambiente físico.....	54
2.2.5 Documentación del sistema.....	55
3 SOFTWARE PARA PROCESAMIENTO DE DATOS	
3.1 Características del <i>software</i>	57
3.1.1 Programas especializados.....	57
3.2 <i>Software</i> por teclado.....	59
3.2.1 IMPS.....	59
3.2.2 CPro.....	72
3.3 <i>Software</i> de reconocimiento óptico.....	78
3.3.1 Evolución del procesamiento de datos.....	78
3.3.2 Descripción de la tecnología.....	80
3.3.3 Diagrama del procesamiento óptico.....	86
3.3.4 Ejemplo de <i>software</i> : Teleform.....	87
3.4 Comparación del software por tecla y óptico.....	93

4 MANEJO DE DATOS Y PRESENTACIÓN DE RESULTADOS

4.1 Sistemas de gestión de datos.....	97
4.1.1 Cobertura y completitud.....	98
4.1.2 Proceso de conversión del dato.....	99
4.1.3 Sistemas de apoyo.....	101
4.2 Presentación de resultados.....	103
4.2.1 Objetivos.....	103
4.2.2 Planificación.....	103
4.2.3 <i>Software</i> para presentación de resultados Redatam..	107
CONCLUSIONES.....	116
RECOMENDACIONES.....	118
BIBLIOGRAFÍA.....	119

ÍNDICE DE ILUSTRACIONES

FIGURAS

1. Menú principal de IMPS.....	61
2. Definición de diccionario de datos en IMPS.....	65
3. Ingreso de datos en IMPS.....	66
4. Pantalla de dibujo de tabulados.....	71
5. Menù de CENTS.....	71
6. Diccionario de datos en CSPro.....	73
7. Ingreso de datos en CSPro.....	75
8. Editor de programas de consistencia.....	77
9. Módulo para cruces rápidos de variables	78
10. Diagrama de flujo para procesamiento óptico.....	86
11. Diagrama de diccionarios escaneados.....	88
12. Creador de formas de Teleform.....	89
13. Estación de escaneo y administrador de los <i>batch</i> escaneados.....	89
14. Administrador de proyecto.....	90
15. Estadísticas de procesos.....	91
16. Definición de tipos de datos de Teleform.....	91
17. Verificación de lectura de Teleform.....	92

18. Sistematización de Redatam+G4.....	12
19. R+xplan y su manejo sencillo.....	113
20. R+xplan y obtención de indicadores.....	113
21. Redatam para la toma de decisiones	
en el plano especial de información geográfica.....	114

GLOSARIO

Cartografía	Dibujo de mapas geográficos
Consistencia	Revisión de la validez de un dato.
Cruces rápidos	Reportes donde se ponen variables en el eje de las "x" y variables en el eje de las "y" para ver su comportamiento.
Diccionario de datos	Descripción de las variables de un archivo, el cual contiene los posibles valores de una variable, la posición inicial dentro del archivo, tipo de dato y largo de la misma.
Diagrama de flujo	Representación del proceso de datos a través de un grafo dirigido.
<i>Fifo</i>	Expresión utilizada para indicar el comportamiento de una cola (primero en entrar, primero en salir).
Imputación	Es la acción de poner un valor controlado o aleatorio a una variable.
Item	Sinónimo de microdato.

Meta datos	Descripción de las características de los datos.
Módulo	Conjunto de programas que se orienta a lograr el mismo fin.
Tabulación	Diseño de cuadros que contienen variables de los cuales se desea ver su comportamiento.
<i>Tunning</i>	Acción de dejar un sistema en su máxima operación.

RESUMEN

El censo es el mayor ejercicio de recopilación y procesamiento de datos que puede existir en un país y que sirve como base para la toma de decisiones tanto en el ámbito público como privado, así como marco de referencia para encuestas de gastos e ingresos, condiciones de vida o cualquier otro tipo de investigación que se desea realizar. Conociendo este hecho es necesario contar con un documento que sirva como base para los Ingenieros en Ciencias y Sistemas y áreas afines para poder desarrollar dicho procesamiento.

Es necesario que las personas que les interesa el procesamiento de datos censal y de encuestas conozcan todos los procesos que se deben llevar a cabo antes, durante y después de procesar los datos.

También es necesario conocer el *software* que existe en el mercado para convertir los datos que provienen del campo a medios digitales, así como el software para validar la información y presentarla al usuario final. Esto es importante ya que este tipo de software es escaso y se mantiene constante durante el tiempo.

OBJETIVOS

Generales

- 1 Desarrollar un estudio del procesamiento de datos masivo para censos y encuestas aplicando tecnología y conocimiento de ingeniería en ciencias y sistemas.
- 2 Generar una guía de procesamiento de datos censales y de encuestas.

Específicos

- 1 Diferenciar los tipos de metodología para procesamiento masivo de datos.
- 2 Determinar los pros y los contras de cada tipo de metodología de procesamiento.
- 3 Encontrar nuevas alternativas basadas en el reconocimiento óptico que beneficien el procesamiento de datos masivos .
- 4 Conocer las diferentes aplicaciones y herramientas que existen para procesar censos y encuestas (en sus diferentes metodologías).

- 5 Conocer los diferentes procesos que se realizan dentro del procesamiento de censos y encuestas.
- 6 Determinar los nuevos adelantos de la ciencia tecnológica aplicados a casos especiales de procesamiento masivo de datos.

INTRODUCCIÒN

Para tomar decisiones políticas, empresariales, económicas y personales bien fundamentadas, las entidades necesitan prever en qué forma dichas decisiones afectarán su entorno. Las respuestas a algunas de las interrogantes más importantes sólo pueden provenir de la información recogida en censos y encuestas. Por ejemplo, ¿cuántos somos?, ¿cómo vivimos?, ¿quiénes son pobres, quiénes son ricos, y por qué? ¿Quiénes utilizan los servicios estatales tales como planteles educativos, clínicas, apoyo agrícola, programas de beneficencia y pensiones de vejez? ¿Es cierto que quienes no aprovechan los servicios públicos pueden acceder a servicios en el sector privado? ¿Quién compra que y por qué? ¿Qué le gusta a alguien y por qué? para responder a estas interrogantes es necesario disponer de datos que abarquen muchos aspectos relacionados con el estudio. En lo que respecta a las encuestas bien fundamentadas y a los censos periódicos en nuestro país, es necesario afirmar que son poco comunes tanto en el ámbito estatal como en las empresas privadas. Sin embargo, debido a la apertura de la inversión local y extranjera poco a poco se necesita un desarrollo constante de la recolección y procesamiento de datos; por lo tanto es necesario capacitar a dirigentes y administradores de proyectos, profesionales en sistemas y analistas para poder procesar la información y brindar rápidamente datos confiables y exactos a quienes lo deseen utilizar.

El estudio de esta tesis está orientado a estudiantes de ingeniería en sistemas y profesionales del área de sistemas para que puedan entender cuáles son los puntos básicos para el procesamiento y entrega de información de los censos y encuestas.

Se pretende que al utilizar esta tesis el profesional de sistemas pueda proporcionar información oportuna, completa, y de alta calidad, sobre la mayoría de los aspectos relacionados. La realización eficiente y eficaz del procesamiento de censos y encuestas puede favorecer que las personas crean en los datos y empiecen a darle la importancia que merecen dichos procesos.

Es importante enfatizar el hecho de que no hay otra labor mayor de procesamiento de datos que un censo de población y habitación, y que cualquier otro censo y encuesta requiere también de capacidad de procesamiento de un alto volumen de información; dado este hecho, es imperativo que los profesionales de sistemas cuenten con una guía para procesar los datos. También incluye una breve descripción de los sistemas que existen para procesar un censo o una encuesta. Es importante hacer notar que los sistemas no son muchos y se mantienen estables dentro del mercado.

1 IMPORTANCIA Y PROPOSITOS DE LOS CENSOS Y ENCUESTAS

1.1 Sistema estadístico integrado

La información sobre la población, sus condiciones de vida y los recursos físicos y materiales de un país es vital para el planeamiento racional de los programas nacionales. Se necesita información precisa y oportuna para tomar decisiones sobre los complejos problemas económicos y sociales que afronta un país. Esto requiere el establecimiento y desarrollo de un sistema estadístico para reunir, coordinar y divulgar la información necesaria. Para países que cuentan con sistemas estadísticos, esto significa expandir el ámbito y mejorar la calidad de los datos proporcionados por el sistema.

Un sistema estadístico no es estático. Evoluciona y crece como función de servicio nacional, dando medidas cuantitativas a través del tiempo. Estas medidas se usan para: (a) medir el progreso o retroceso en el logro de metas nacionales y (b) evaluar la efectividad de programas y políticas relacionadas con el desarrollo y bienestar del país.

Durante los últimos años, ha habido un interés creciente por las estadísticas como medio para orientar muchos aspectos del desarrollo de un país. Estos incluyen las funciones de gobierno a todo nivel, el crecimiento y complejidad creciente de la industria y los negocios y el impacto resultante de cambios en la sociedad.

En consecuencia, los gobiernos están más conscientes de su responsabilidad de proporcionar servicios estadísticos nacionales.

Cuando se desarrollan diversas clases y series estadísticas, se hace evidente que hay necesidad de una coordinación y una colaboración para su evolución y mantenimiento con el fin de evitar la duplicación innecesaria de esfuerzos e inconsistencias en el resultado. Hay también preocupación por mejorar la integridad y calidad del producto estadístico. Así, a través del tiempo, existe un creciente interés por lograr un sistema estadístico integrado a nivel nacional. Los países procuran alcanzar esta meta de diferentes maneras y a través de diversas estructuras organizativas.

Esta sección analiza los objetivos y factores principales para desarrollar un sistema estadístico integrado, denominado de aquí en adelante simplemente como sistema integrado. La sección también describe sistemas centralizados y descentralizados, que son los principales estructuras organizativas. Debe observarse que un sistema integrado no funcionará bien a menos que la coordinación de las actividades estadísticas del país se concentren en alguna oficina o institución que esté claramente autorizada y se le haya asignado esa responsabilidad. Sea cual fuere la organización existente, la integración incumbe al programa total de estadísticas oficiales del país.

1.1.1 Objetivos de un sistema integrado

El objetivo básico de cualquier sistema estadístico es recolectar, procesar y publicar diversas series estadísticas e intervalos regulares; las estadísticas deben proporcionar información cuantitativa útil para satisfacer tanto los intereses gubernamentales como los no gubernamentales.

Para servir a sus usuarios en la forma más eficiente, el sistema debe establecer ciertas prioridades para los datos que proporciona. Por ejemplo, las estadísticas de producción de la industria, bancos, comercio exterior y similares pueden no ser apropiadas en un país en desarrollo donde la economía está basada en actividades agrícolas menores y artesanías caseras en los que la unidad de trabajo es la familia y la mayor parte del intercambio es por trueque. En los países donde se está desarrollando un sistema integrado, se deben reconocer cinco amplios objetivos:

- 1) Expandir los datos básicos para medir el estatus de los programas nacionales.
- 2) Mejorar la calidad y oportunidad de las estadísticas que se están produciendo.
- 3) Dar mayor importancia y concentrarse en forma más efectiva en la necesidad de estadísticas demográficas y sociales.
- 4) Identificar vacíos significativos en las necesidades de información.
- 5) Mejorar la utilidad de las estadísticas tanto para el gobierno como para los sectores no gubernamentales del país.

1.1.2 Recolección y publicación de datos

Si un sistema integrado va a ser efectivo, los temas para los cuales se recolecten datos deben reflejar la naturaleza de la estructura social del país, su economía, su etapa de desarrollo y la estructura de sus instituciones.

Esto requiere un enfoque objetivo en lo que se refiere a la recolección y publicación de datos estadísticos que surjan de un plan maestro para un sistema integrado. Para ello se requieren métodos con los cuales se puedan procesar los datos de manera inmediata y se pueda presentar la información oportuna.

1.1.2.1 Desarrollo de normas estadísticas

Desde su inicio, un sistema integrado necesita preocuparse por la calidad de los datos que recolecta y publica. La calidad se alcanza estableciendo y reforzando normas estadísticas. En general, hay tres elementos principales que deben incluir las normas estadísticas:

- 1) Definición precisa de cada *ítem* de una serie estadística determinada.
- 2) Agrupación sistemática de estos *ítems* en una clasificación lógica.
- 3) Técnicas o métodos específicos que se deban seguir al construir la serie estadística y controlar su calidad.

La documentación de estos elementos, incluyendo procedimientos e instrucciones fácilmente entendibles, es esencial para la formulación y aplicación de normas estadísticas. Para ciertas áreas temáticas, Naciones Unidas ha desarrollado guías respecto a conceptos, clasificaciones y métodos.

a) Conceptos

Los datos estadísticos pueden conducir a confusión y a error a menos que haya definiciones claras de los *ítems* que se presentan. Es más, los conceptos (y definiciones) tienen que uniformarse de modo que un *ítem* tenga el mismo significado cuando aparece en varios cuadros de una serie o cuando se usa en varias series estadísticas.

b) Clasificaciones

Las clasificaciones se usan en casi todos los campos de la estadística. La clasificación es la única forma de resumir las observaciones individuales o los *items* de datos de manera que se reproduzcan cifras significativas. Involucra la ordenación sistemática de *items* de datos en grupo o categorías de acuerdo con criterios establecidos lógicamente. La clasificación uniforme es el único método por el cual los elementos comunes o asociados de dos conjuntos de datos diferentes pueden compararse. Por ejemplo, una clasificación uniforme de la actividad económica permite la comparación entre censos de población y censos agropecuarios o industriales.

c) Metodología

Para alcanzar calidad y consistencia de los datos estadísticos, es esencial usar métodos estadísticos uniformes. Este enfoque permite comparar las estadísticas entre áreas y períodos.

La aplicación de métodos estadísticos uniformes en la recolección, procesamiento, presentación y análisis de datos mejorará la calidad de las estadísticas emitidas por cualquier sistema estadístico.

d) Desarrollo de la investigación en métodos estadísticos

Puede aumentar la calidad de los datos estadísticos y disminuirse substancialmente los costos, desarrollando y probando métodos estadísticos nuevos para la resolución y procesamiento de datos. Muchos países han establecido grupos separados de personal dentro de sus sistemas estadísticos para hacer trabajos de desarrollo e investigación en metodología estadística aplicada. Los resultados y los beneficios substanciales de sus esfuerzos pueden encontrarse en la literatura estadística.

Muchos de los problemas prácticos encontrados en la recolección y procesamiento de datos pueden resolverse por medio de esa investigación y con ella aumentar el potencial para el mejoramiento en la calidad del producto.

1.1.2.2 Coordinación para actividades de encuestas

La coordinación hace posible la producción de estadísticas bien integradas de la manera más eficiente y económica. Puede eliminar duplicación innecesaria de esfuerzos de instituciones públicas que recolectan y publican datos estadísticos y evitar operaciones no esenciales. Una autoridad coordinadora puede ser también efectiva para mejorar la calidad de las estadísticas insistiendo en la aplicación de normas y procedimientos adecuados entre los diversos productores de datos.

En un país en desarrollo, donde tanto los recursos financieros como el personal estadístico profesional pueden ser extremadamente limitados, una autoridad estadística coordinadora puede ser especialmente valiosa.

Tal autoridad no sólo puede establecer los objetivos y prioridades estadísticas aplicables a las necesidades del país, sino que también puede ser monitora de programas estadísticos que se establezcan para responder a tales necesidades. Una autoridad coordinadora responsable es particularmente importante en un país donde las actividades estadísticas no estén centralizadas.

1.1.2.3 Coordinación de los censos

En algunos países, se levantan dos censos simultáneos o casi juntos porque hay elementos esenciales comunes a ambos. Por ejemplo, los censos de población y habitación se realizan conjuntamente o como actividades separadas pero bien coordinadas. Esto permite obtener información sobre las características personales de los miembros del hogar para relacionarlos más fácilmente con datos sobre sus condiciones de vida; además, se facilita el procesamiento de los resultados.

Los censos de población y habitación pueden usarse en la preparación de otros censos, siempre que se tomen las providencias en el cuestionario.

Se podría agregar una pregunta sobre actividad agropecuaria de modo que la oficina de estadística pudiera ajustar áreas de empadronamiento para los propósitos del censo agropecuario.

Otra posibilidad es incluir todas las estructuras no residenciales en el listado para el censo de población de modo que las unidades no residenciales puedan ser identificadas para un censo económico. Este procedimiento debería considerarse muy cuidadosamente.

Los empadronadores de censo de población están instruidos para ubicar personas y su lugar de habitación; listar estructuras no residenciales puede llevar a confusión y puede afectar adversamente la calidad de empadronamiento.

Otros censos no pueden ser manejados en conjunto tan ventajosamente como los de población y habitación, porque sus elementos esenciales son diferentes. Por ejemplo, la unidad de empadronamiento para un censo agropecuario es la explotación agrícola, lo que difiere del hogar o la unidad de vivienda usada en censos de población y habitación.

No obstante, algunos países han efectuado estos censos conjuntamente para ahorrar tiempo en viajes y obtener datos de población para relacionarlos con la información agropecuaria. Los censos económicos, además, usan una unidad de empadronamiento enteramente diferente: el establecimiento. No es práctico combinar censos económicos con censos de población, habitación o agropecuarios.

Al tomar una decisión de integrar las actividades de recolección de datos, el personal estadístico debe considerar la carga adicional para el informante si se combinan varias encuestas, lo mismo que la potencial confusión que puede resultar para el empadronador. Si los censos se combinan, sería beneficioso usar los mismos conceptos y procedimientos de la manera que correspondan.

1.1.3 Elementos de un sistema estadístico integrado

Para que un sistema estadístico integrado tenga éxito debe poseer varios elementos básicos. Los elementos deben combinarse para establecer un programa integrado total con objetivos tanto a corto como a largo plazo.

Un buen sistema ayuda en la distribución adecuada de los recursos para hacer frente a estos objetivos. Un elemento importante es revisar los cuestionarios y encuestas para evitar duplicación de esfuerzos. Los encargados de la revisión pueden mejorar la utilidad de los datos que están recolectando por medio de una selección cuidadosa de las preguntas que se van a efectuar, sugiriendo diseños de tabulación que respondan a los diferentes requerimientos o consolidando dos o más encuestas en una.

1.1.4 Factores para desarrollar un sistema integrado

Hay muchos factores que influyen en la clase de sistema integrado que establezca un país.

Mientras que el sistema que eventualmente se adopte puede no ser ideal y puede que ni siquiera represente el mejor sistema estadístico que se acomode a todas las necesidades de información, su objetivo principal es producir un conjunto de datos estadísticos completo e integrado.

1.1.4.1 Niveles de coordinación

El proceso de coordinación es manejado o puede ser manejado por un grupo de personas, ya sea que pertenezcan a entidades estadísticas o consejos independientes que actúen como asesores o sean empleados de gobierno a quienes se les asigne esta responsabilidad.

Los esfuerzos de coordinación tienen que ver con la utilización máxima de los recursos, tanto humanos como de equipo; la planificación y manutención real del calendario; los controles de costos; el cumplimiento de las prioridades, y la solución de desacuerdos.

La coordinación es esencial en todos los aspectos, incluyendo las actividades de planificación, recolección, procesamiento y divulgación del sistema estadístico. La cantidad o grado de coordinación puede ejercerse a diversos niveles dentro de estas actividades.

En forma similar, la formulación y ejecución de programas de computador generalizados debieran ser coordinadas por especialistas en computación y en el tema específico.

Esto llevará a efectuar sistemas especializados que cumplan con los requerimientos de los censos y encuestas, y evitará el esfuerzo Independiente de los especialistas por desarrollar programas diferentes según sea la temática de los censos o encuestas.

1.1.4.2 Integración de la recolección de datos

La mayor ventaja de un sistema integrado es la consolidación de varios requerimientos de datos en una sola encuesta o censo, siempre que las unidades de empadronamiento sean compatibles. Tal consolidación reduce significativamente la carga total para el informante. En repetidas ocasiones se ha demostrado que los mayores factores de costo en el terreno son transporte, ubicación del informante y explicación de la razón de la investigación. La entrevista misma representa un costo menor y se puede manejar una cantidad razonable de *items* adicionales con poco gasto adicional.

Otra ventaja de la recolección integrada de datos es la mayor probabilidad de contar con entrevistadores experimentados. También hay mayor probabilidad de supervisión efectiva del empadronamiento puesto que el personal de terreno conocerá mejor una gran variedad de temas.

La recolección integrada de datos permite la aplicación de definiciones uniformes y la consistencia en la manera en que los entrevistadores hagan las preguntas. Permite también la tabulación cruzada, el análisis y la correlación de características adicionales que son imposibles cuando se trabaja con encuestas separadas.

1.1.4.3 Integración del procesamiento de datos y publicación

Con el uso del computador, se puede hacer mucho para integrar el procesamiento y la publicación de los datos. Pueden diseñarse programas de modo que los sistemas de computador electrónico impriman cuadros estadísticos en forma definitiva.

Existen muchas ventajas al hacer que el computador prepare los cuadros listos para la publicación, pero la disponibilidad limitada de personas con experiencia en métodos de tabulación, programación en computador y análisis de sistemas convierte en un imperativo la integración del procesamiento de datos.

Integrando la publicación de datos de diversas fuentes, algunos miembros del personal pueden llegar a ser especialistas en presentación de datos. Esto trae como consecuencia mejores publicaciones y más oportunas. Además, promueve una mejor planificación de la publicación, la normalización de su presentación y ayuda a hacer más expedita la impresión y distribución de las publicaciones.

1.1.4.4 Precauciones

Al considerar un sistema integrado de datos, los planificadores tienen que tener cuidado de no combinar demasiados proyectos en una actividad estadística o intentar fusionar diversos programas que impliquen complicaciones para el análisis o la recolección de datos. Por ejemplo, combinar una encuesta sobre las características de la población con un estudio de establecimientos industriales puede conducir a errores de respuesta, errores de medición y errores de interpretación de los resultados.

1.1.5 Responsabilidad de un sistema integrado

Cuando se desarrolla un sistema integrado para proporcionar servicios estadísticos en un país, éste asume serias responsabilidades que pueden ser cruciales para el bienestar y el desarrollo del país. Por esta razón es muy importante que la responsabilidad de coordinación se ubique en uno de los ministerios más importantes del gobierno, tales como el Ministerio de economía. Puesto que el sistema constituye un canal de información de doble vía entre informantes y usuarios de datos, tiene que equilibrar la necesidad y el uso esperado de las estadísticas con los problemas de confiabilidad, moderación de la información deseada y calidad esperada de los datos producidos.

Un sistema integrado debe estar consciente de las prioridades de información y adherirse a ellas por medio de un estrecho contacto y cooperación con los diversos ministerios del gobierno. Las personas responsables deben evaluar las ventajas y desventajas de combinar diversos estudios al recoger los datos. Puede producirse una fuerte reducción de costos con tales combinaciones, lo que debe sopesarse con la carga que significa para el informante.

Debe tenerse en cuenta la oportunidad en la publicación de los resultados y la diferencia potencial en la calidad de los datos de un cuestionario que cubra varios temas. Ejemplos típicos son las consideraciones al cubrir tanto las investigaciones de población y habitación en un solo censo o combinar el empadronamiento 100 por ciento simultáneo y hacer preguntas adicionales a una muestra de la población. Esto tiene que evaluarse.

Otra decisión que debe tomarse es el manejo de algunos *items* de datos por medio de una investigación de seguimiento por muestra, sea en conjunto con un control de evaluación del censo o independientemente para proporcionar mayor detalle debajo del nivel nacional.

1.1.5.1 Control de calidad

Ya que se enfrenta una gran necesidad de datos y usualmente se opera con fondos muy limitados, un sistema estadístico tiene tendencia natural a maximizar la producción de datos y rebajar su nivel de calidad. Es responsabilidad del sistema estadístico nacional desarrollar y mantener programas que aseguren la integridad de los datos que produce.

Estos incluyen estudios de evaluación tanto de la cobertura como de las características del censo. La recolección de datos debe ser controlada, el procesamiento de los resultados debe ser verificado y los programas de computador deben incluir técnicas de auto-control y límites razonables para asegurarse de que las tabulaciones produzcan el conjunto de características de datos exigido en las especificaciones.

1.1.5.2 Sensibilidad del sistema

Un sistema estadístico integrado es idealmente sensible a los requerimientos e intereses de los usuarios que lo producen. Para funcionar efectivamente, el sistema debe buscar regularmente el consejo y las recomendaciones de aquellos que necesitan y usan los datos que proporciona.

Esto puede hacerse informalmente por medio de encuestas, por muestreo de los usuarios, en las cuales se busque medir el valor de las estadísticas. Tal análisis ayudaría a determinar prioridades en lo que se refiere a qué datos se recogerán y publicarán o se pondrán a disposición como tabulación especial.

El sistema también ayuda a determinar cuáles *items* de datos deben retenerse para información de las tendencias históricas y qué nuevas clases de datos se necesitan para medir los actuales aspectos del desarrollo y los requerimientos de información del país. Mantener buenas relaciones con los usuarios y tenerles información disponible oportunamente puede mejorar muchísimo su deseo de ayudar a hacer que el público responda a los censos y encuestas por muestreo. Este esfuerzo, de hecho, puede llevar a contribuciones efectivas para el funcionamiento del sistema de información estadística.

1.1.6 Integración de censos y encuestas por muestreo

Los estadísticos en la mayoría de los países reconocen que un censo proporciona sólo algunos de los datos que necesitan los encargados de la toma de decisiones. Sobre una variedad de temas se necesitaría producir datos más detallados o especializados durante el período intercensal.

Para máxima utilidad, los datos recogidos en una encuesta durante el período intercensal deben integrarse con los datos del censo, lo mismo que los de otras encuestas intercensales y aun de registros administrativos. Además, las unidades de medición y los conceptos deben ser consistentes, de modo que los resultados de varias encuestas puedan ser analizados y relacionados. Por ejemplo, el censo podría recoger los datos básicos sobre fecundidad y nivel educacional; las encuestas intercensales podrían aumentar los detalles para incluir fecha y sexo de cada nacimiento y campo de educación.

Para esta integración es necesario crear sistemas computacionales que den mantenimiento a las diferentes variables estudiadas con posibilidad de comparatividad en el tiempo y con agrupación en unidades geográficas de análisis.

1.2 Papel y objetivos de los censos y encuestas

1.2.1 Papel de los censos y encuestas

El papel o función de un censo es inventariar los recursos y características de un país en un momento dado. El inventario puede relacionarse con las características de la población; su vivienda; sus actividades económicas, como la agricultura, los negocios, la industria o el comercio; o casi cualquier condición física o no física que afecte el bienestar de los habitantes del país.

Los censos se originaron hace muchos siglos en el empadronamiento de la población en áreas específicas con propósitos de administración, tributación, reclutamiento y muchos tipos de planificación.

Aún cuando los problemas que se enfrentan al realizar un censo no son nuevos, las necesidades de datos a las cuales apuntan los censos modernos se han expandido muchísimo, aumentando con eso el alcance y la complejidad del esfuerzo censal.

El número de habitantes del mundo está aumentando con una tasa que apunta a un incremento substancial de la población.

1.2.1.1 Papel para responder a los requerimientos estadísticos

Para planificar y ejecutar programas y actividades, los funcionarios del gobierno a diversos niveles y los usuarios particulares, como el comercio, la industria, los organismos dedicados a la investigación y el público en general necesitan estadísticas.

Esas estadísticas pueden clasificarse en dos categorías:

La primera categoría puede obtenerse de censos completos, que proporcionan datos por áreas pequeñas, además de resúmenes nacionales. Típicamente los censos se realizan en períodos preestablecidos, generalmente en intervalos de 10 años, para proporcionar información que permita medir cambios o efectuar comparaciones con datos anteriores. Un censo también sirve como base de datos para otras características que puedan recogerse por primera vez.

La segunda categoría corresponde a datos requeridos con mayor frecuencia; las estadísticas se recopilan mediante encuestas continuas por muestreo para medir cambios a corto plazo en todo el país y en unas pocas áreas mayores.

Un censo, a pesar de su periodicidad, debe considerarse parte del programa de estadísticas continuas de un país. Es necesario mantener un equipo profesional pequeño para evaluar la utilidad del censo pasado y considerar mejoras que puedan incorporarse al censo siguiente.

Los datos del censo son básicos para muchas decisiones concernientes a representación política a todos los niveles de gobierno. Se usan para la asignación de fondos y la ubicación de servicios públicos basados en la distribución de la población y proporcionan información esencial para establecer qué planeamiento futuro y qué acciones administrativas pueden determinarse. Además, los datos del censo se usan para complementar el programa de encuestas por muestreo durante el período intercensal.

1.2.1.2 Metas generales del programa censal

Las metas generales de unos censos de población y habitación son reunir y divulgar información sobre características demográficas, sociales, económicas y habitacionales de la población. Una meta principal en este sentido es alcanzar precisión en los datos y oportunidad para tenerlos disponibles para el uso. Esto es particularmente importante debido a que una preocupación fundamental de cada censo nacional debe ser proporcionar datos como base para la formulación de políticas de gobierno y para ejecutar programas y acciones.

1.2.2 Objetivos estadísticos del censo

Un censo de población y habitación es un esfuerzo estadístico mayor que involucra una inversión considerable de los recursos públicos de un país.

Es esencial, por lo tanto, que los objetivos del censo se revisen totalmente, se aclaren para todos y cada uno de los involucrados en su realización y se les dé amplia circulación tanto entre los informantes como entre los usuarios potenciales.

Puesto que los datos obtenidos en el censo servirán para muchos propósitos, es crucial determinar sus objetivos antes de dar término a los cuestionarios y dar comienzo a la recolección de datos. Los cambios posteriores de propósitos o de objetivos pueden producir atrasos costosos y posibles deterioros de la calidad de datos.

1.2.2.1 Medir las condiciones existentes

El primer objetivo es proporcionar datos sobre las condiciones económicas y sociales existentes en el país. Si ha habido un censo previo, el análisis de los cambios indicados por el último censo puede revelar el crecimiento de la población, los cambios significativos en la distribución por edades, los sectores donde exista hacinamiento de vivienda y las áreas económicas deficitarias.

Puede también indicar si el nivel de educación ha mejorado y cuánto, lo mismo que muchas otras condiciones que identifiquen los problemas o las localidades donde el gobierno debe concentrar sus esfuerzos para lograr el progreso.

El censo, que se efectúa en intervalos relativamente poco frecuentes, mide tendencias a largo plazo. Siempre que sea posible, es útil mantener la comparabilidad entre censos en las preguntas que se hacen, las definiciones usadas y los procedimientos que se aplican en la recolección y análisis de los datos; sin embargo, la importancia de un *item* y su concepto no deben sacrificarse en aras de la comparabilidad.

1.2.2.2 Planificar programas de desarrollo

Los censos de población y habitación proporcionan mediciones sociales y económicas que se usan como base para la formulación y el planeamiento de la política gubernamental.

La distribución y las características de la población y sus viviendas son esenciales para un examen legislativo informado de los programas de desarrollo económico y social destinados a mejorar las posibilidades de empleo, salud pública y bienestar.

Un censo proporciona importantes datos que son la base para efectuar mediciones del PIB (producto interno bruto) de un país, el cual es un índice útil para medir la estabilidad y el progreso económico.

1.2.2.3 Establecer las bases para ejercer la acción

Las estadísticas censales las usan comúnmente las instituciones de gobierno, empresas comerciales y particulares como base para actuar en el sentido de corregir, mejorar o expandir ciertas funciones.

Por ejemplo, el gobierno podría desear establecer un programa de educación para adultos en aquellas áreas donde los resultados del censo mostraran una alta tasa de analfabetismo de adultos, o establecer programas públicos de vivienda en áreas donde haya evidencia de hacinamiento y deterioro. La divulgación de la información censal es esencial si todos los usuarios potenciales van a estar conscientes de su valor como base para tomar decisiones o ejercer ciertas acciones.

1.2.2.4 Proporcionar datos para áreas pequeñas

Con la creciente urbanización y las mejores métodos de comunicación en muchos países, ha habido un interés cada vez mayor por los datos censales para áreas pequeñas dentro de las ciudades y de otras divisiones administrativas. Estos incluyen policía local y jurisdicciones de protección de incendios, distritos escolares, áreas de salud pública, áreas de planeamiento urbano de muchas clases, zonas para fines comerciales, zonas de tránsito para calcular cargas y necesidades de transporte y muchos otros tipos de sectores geográficamente identificables que son parte de un complejo mayor. Debido a los costos involucrados en la recolección de datos, un censo es usualmente la única fuente viable de buenos datos para áreas pequeñas como éstas.

Las demandas de datos para áreas pequeñas requieren de una mayor flexibilidad en disponibilidad de datos que pueden no encontrarse en la serie de publicaciones regulares de los resultados del censo. Un objetivo importante del programa censal es atender a tales necesidades por medio de tabulaciones especiales y agregaciones específicas para áreas estadísticas muy pequeñas, como áreas censales, manzanas u otras subunidades fácilmente identificables. Esto permite al usuario examinar los datos de áreas muy pequeñas y recombinarlas en áreas mayores según convenga.

1.2.2.5 Guiar la investigación

Además de los usos legales, administrativos y de programa específico, los censos proporcionan datos de vital importancia para un análisis y evaluación detallados de la composición, distribución, crecimiento y cambios en la población y la vivienda.

Preguntas de interés demográfico e investigaciones tales como los cambios en la distribución urbano-rural de la población y su distribución geográfica de acuerdo con variables, como ocupación, industria, nivel de educación y edad, son respondidas por los resultados del censo. Asimismo, la disponibilidad de datos por áreas pequeñas permite al investigador identificar temas de análisis y reagrupar las áreas de diversos modos según los objetivos de la investigación.

1.2.3 Objetivos operativos del censo y encuestas

Producir las estadísticas requeridas es ciertamente la razón primera para levantar un censo. No obstante, un censo tiene varios objetivos operacionales que son importantes para el programa estadístico de un país.

1.2.3.1 Establecer un marco para encuestas intercensales

Los censos de población y habitación se han hecho cada vez más importantes como marcos para el desarrollo de otros procedimientos y sistemas de recolección de datos. Ellos proporcionan la base de datos para la elaboración de diseños muestrales confiables que pueden usarse para efectuar encuestas intercensales eficientes sobre tópicos para los cuales es importante la observación frecuente; de esta manera es trascendente que las tabulaciones por áreas de empadronamiento, mapas y otros documentos pertinentes se conserven adecuadamente. Además, los datos censales proporcionan la base sobre la cual se pueden construir diversas series estadísticas; pueden usarse también para verificar la precisión y mejorar la calidad y comparabilidad de las series de datos existentes.

En forma alternativa, los resultados del censo pueden combinarse con diversos registros administrativos para desarrollar estimaciones actualizadas y confiables para muchos programas de gobierno que tienen que ver con pensiones, requerimientos en educación, seguro social, salud pública y otros servicios deseables.

1.2.3.2 Servir de prueba para censos futuros

Un subproducto de un censo es la oportunidad de probar ciertos procedimientos bajo condiciones reales para mejorar censos futuros. Sin embargo, este objetivo debe restringirse para evitar atrasos excepcionales y posibles costos adicionales. Hay varias pruebas potencialmente útiles que pueden realizarse a escala muy pequeña y que pueden ser muy útiles para planificar censos futuros. Ellas incluyen el estudio de la carga de trabajo de diferentes empadronadores, de métodos alternativos de adiestramiento, de procedimientos de codificación y de procedimientos de revisión de tabulaciones similares.

1.2.4 Consideraciones al ejecutar un censo o encuesta

Pueden surgir grandes problemas en la realización del censo debido a la variedad de propósitos que se deben cumplir. El dilema es cómo organizar y cumplir el programa del censo, para maximizar su utilidad, con recursos limitados. Los siguientes elementos deben recibir atención constante.

1.2.4.1 Procesamiento de datos oportuno

Con demasiada frecuencia un censo cae en desuso, y el esfuerzo y el costo se desperdician porque se ha demorado demasiado en tener disponibles los datos. Para que sean de máximo valor, los resultados deben tabularse y publicarse lo más rápido que sea posible.

Muchos países siguen la práctica de emitir tabulaciones preliminares de características básicas, como la población por divisiones administrativas por sexo y edad, pocos meses después que se haya completado la recolección de datos. Las tabulaciones preliminares se amplían con informes que contienen distribuciones simples y tabulaciones cruzadas significativas de las características; éstas se emiten varios meses después de completado el procedimiento, pero antes de que estén listas en forma de publicación final.

Este procedimiento trata de responder a la crítica común de que la información es demasiado antigua para tener valor actual cuando no está disponible hasta varios años después de efectuado el censo.

1.2.4.2 Oportunidad de los datos

Los recursos esenciales para que un censo tenga éxito pueden resumirse en cuatro palabras: dinero, personal, conocimiento y máquinas. Si estos recursos están disponibles y cada uno se usa tratando de lograr máxima ventaja, hay grandes probabilidades de que el programa del censo tenga éxito.

Al mismo tiempo, los recursos estadísticos disponibles deben examinarse para eliminar duplicaciones.

Estos recursos son de muy diverso tipo, e incluyen registros administrativos de oficinas de gobierno, tanto nacionales como locales, que pueden contener la información requerida si se procesan y resumen adecuadamente. Si los registros administrativos de un programa de gobierno en operación proporcionan antecedentes estadísticos para un segmento medible de la población, puede no ser económico o factible duplicar la obtención de tal información en un censo. Dicha información puede ser suficientemente útil como para permitir la asignación de recursos censales a otros temas necesarios para los cuales de otro modo hay muy poca o ninguna información disponible.

El uso juicioso de los recursos disponibles involucra también un inventario de las informaciones necesarias que pueden obtenerse adecuadamente por medio de un censo. Esto puede involucrar consideraciones para ampliar, limitar o modificar el programa censal con el fin de producir datos que sean del mayor beneficio para el desarrollo total económico y social del país.

1.2.4.3.1 Responsabilidad por el censo

Normalmente un levantamiento censal es responsabilidad de la oficina nacional de estadística. La tarea específica de planificar y dirigir las múltiples operaciones que se requieren será asignada a personal de estadísticos con preparación profesional.

La centralización logra uniformidad en el método para realizar el censo y compilar los resultados, lo mismo que consistencia en los conceptos y en las definiciones. Igualmente importantes son los beneficios en eficiencia de operaciones y disminución de costos que puede lograr la realización centralizada del censo.

En varios países, la oficina nacional de estadísticas consigue la cooperación del gobierno local en diversos niveles para la operación de recolección de datos, puesto que se requieren temporalmente un gran número de empadronadores. En otros países, el personal necesario para efectuar el empadronamiento lo proporcionan otros componentes del gobierno nacional, como el servicio de correos. Si el sistema estadístico de un país es descentralizado, es necesario que el ministerio que tenga la responsabilidad del censo designe una oficina o institución específica que actúe como organismo principal para dirigir el programa.

1.2.4.4 Tipos de mediciones estadísticas requeridas

El objetivo primario de un censo es proporcionar estadísticas útiles a diversos usuarios de datos. Habitualmente se necesitan mediciones de la población total, desempleo, nivel de educación, condición de la vivienda y otros similares para orientar el planeamiento, la investigación y la administración de programas sociales, económicos, de salud y otros.

Los siguientes ejemplos son de la clase de preguntas que deberían responderse antes de recomendar la inclusión de un tópico en el censo:

- 1) ¿Hasta qué punto está bien definido el concepto que se va a medir?
¿Hay clara comprensión de lo que quiere decir desempleo, alfabetismo, hacinamiento, vivienda inadecuada, ingreso total de la familia, etc.?
- 2) ¿Cuánto detalle es necesario y qué nivel de precisión se necesita para diversos niveles de detalle? Estas consideraciones tienen un impacto substancial sobre los costos de un censo.
- 3) ¿Qué frecuencia de observación es esencial o significativa? Muchos tópicos dinámicos demandan una medición más frecuente que la proporcionada por un ciclo censal de 5 ó 10 años; otros tópicos pueden ser importantes en un período determinado, pero no necesitan medición continua a través del tiempo. Investigaciones como éstas pueden ser mejor atendidas con encuestas por muestreo que ocasionan menos costos.
- 4) ¿Cuánta estabilidad tienen los fenómenos que se van a observar? ¿Es el momento de la medición un aspecto crucial para la confiabilidad de los datos? ¿Es un censo periódico el instrumento adecuado cuando es difícil obtener una buena respuesta?
- 5) ¿Qué sistema de clasificación aplicado a estos datos ayudará a incrementar su utilidad? ¿Sería, por ejemplo, adecuada para el país la clasificación ocupacional estándar?

- 6) ¿Qué forma de salida de datos contribuirá a mejorar su utilidad? ¿Servirían para este propósito las tablas estadísticas? ¿Los diagramas y los gráficos de ciertas relaciones claves facilitarían a los usuarios la comprensión de los datos? ¿Serían las medianas o los promedios de más valor que las distribuciones porcentuales?

1.2.5 La precisión, un objetivo del censo y encuesta

Al considerar las necesidades de información, un aspecto que rara vez recibe atención suficiente es el de los requerimientos de precisión y la importancia de poder confiar en los resultados.

La información que se usa para orientar las decisiones respecto a importantes programas y problemas debe responder a ciertas normas. La precisión de la información clave debe ser medible y debe estar en un nivel aceptable. En general, la precisión debe ser suficiente como para evitar la interpretación falsa que puede originarse como resultado de mediciones no confiables o seriamente sesgadas.

1.2.5.1 Calidad y confiabilidad

Las personas que planeen y diseñen programas censales deben estar conscientes de las fuentes de error, tanto muestrales como no muestrales. Una base para medir el error asociado al muestreo puede lograrse por medio de la aplicación de técnicas de muestreo.

Al error no muestral debe presentársele aún mayor atención, ya que puede ser lo bastante grande como para invalidar la utilidad de los resultados.

Controlar el error muestral requiere cuidadosa atención a temas como los siguientes:

- a. Asegurarse que la redacción de cada *item* del cuestionario sea sencilla y breve, en términos que sean comprensibles para el informante.
- b. Hacer preguntas que pueda responder el informante o evitar preguntas que estén sujetas a excesivo sesgo por memoria, a menos que el informante mantenga registros regulares.
- c. Asegurarse de que el cuestionario sea de un tamaño que no produzca la fatiga o falta de atención del informante de modo que afecte la exactitud de la respuesta.
- d. Efectuar pruebas de los planes censales con un pequeño grupo de informantes, para ayudar a definir procedimientos y preguntas que satisfagan los objetivos.
- e. Proporcionar adiestramiento y supervisión adecuados para la operación de recolección de datos.
- f. Revisar los procedimientos de procesamiento de datos y los resultados tabulados.

1.2.5.2 Evaluación de los resultados

Es conveniente para el país incluir alguna actividad que evalúe tanto las estadísticas producidas como los procedimientos usados en la realización del censo.

La evaluación de las estadísticas orientará a los usuarios en la interpretación y aplicación adecuada de los resultados; la evaluación de procedimientos será beneficiosa principalmente para el planeamiento de censos y futuros programas estadísticos. Como mínimo, es conveniente medir el grado de cobertura de la población total y de subgrupos específicos, puesto que muchas decisiones se basan en estadísticas de población.

1.2.6 Importancia de actualizar los tópicos

Esfuerzos sucesivos y modificaciones para servir nuevas necesidades a medida que ellas se desarrollan constituyen un enfoque práctico y lógico para lograr un sistema útil de estadísticas oficiales.

De este modo, el diseño de un sistema de información general como es un censo debe tener carácter evolutivo, identificando nuevas necesidades o tomando ventaja de nuevos recursos y construyendo sobre lo que ya hay. La evolución de fuentes de datos, tanto existentes como nuevas, debe ser continua. Estas fuentes deben modificarse, eliminarse o ampliarse a medida que el país aprende a mejorar el sistema y a medida que las necesidades anteriores cambian o desaparecen y surgen nuevas necesidades.

Es esencial prestar atención constante a la utilidad de los datos recolectados y a las diversas tabulaciones planificadas y publicadas. Los administradores del censo deben determinar si se está entregando a los usuarios lo que les interesa o, más importante, lo que requieren para nuevos programas de gobierno en desarrollo o decisiones comerciales y mercantiles.

Por ejemplo, la pregunta sobre asistencia escolar puede haberse hecho en censos anteriores sólo a personas de hasta 12 años de edad.

A medida que la educación se hace más universal y más personas reciben educación superior al nivel primario y secundario, la pregunta debe cubrir en censos posteriores más grupos de edad. En forma similar, con el interés creciente por el tipo de educación, se espera que la encuesta sobre nivel alcanzado reconozca el adiestramiento técnico además de los grados académicos universitarios y de postgrado.

Otros ejemplos de los cambios en las investigaciones censales para medir las condiciones que se consideran generalmente importantes son encuestas en campos de estudio especializados, como mediciones de capacitación laboral y de reserva de fuerza de trabajo.

Con la creciente movilidad y migración interna, particularmente a las áreas urbanas, serían útiles las preguntas sobre lugar de residencia previa y fecha de traslado. Nuevos tópicos para medir el nivel de vida pueden incluir *items* tales como la disponibilidad de agua caliente, refrigeración y calefacción central o aire acondicionado en la vivienda.

1.3 Etapas preliminares al procesamiento

Para llevar a cabo el procesamiento de datos de un censo y encuestas es necesario elaborar puntos previos los cuales llevarán a tener la información recolectada y centralizada para su posterior proceso. A continuación se detallan pasos previos sin incluir todo lo que tiene que ver con administración, contratación de personal, etc.:

1.3.1 Planificación y presupuestos

Tiene que ver con la elaboración de planes detallados en donde se establecen fechas límites para cada paso, así como la elaboración de presupuestos que incluyen materiales, equipos, recursos humanos, recursos administrativos y otros.

1.3.2. Definición de temática por investigar

Se debe de establecer qué es lo que se desea investigar y luego desarrollar un cuestionario con manuales para que sean llenados correctamente.

El diseño del cuestionario tiene que ser orientado a facilitar el manejo en campo por parte de los empadronadores o encuestadores. Este diseño no debe pasar por alto el procesamiento de datos posterior, ya que debe ser orientado a un fácil ingreso o escaneo del mismo, según sea el caso.

1.3.3 Actualización cartográfica

Para un censo se debe realizar un recorrido en la totalidad del territorio para establecer lugares poblados con viviendas y características generales. Producto de la actualización cartográfica, se debe tener una base de datos, la cual sirve como parámetro de cobertura al elaborar un censo.

Además, la actualización cartográfica debe de dar como producto un archivo base de mapas, planos y croquis con la representación de todos los lugares poblados que se encuentren. Con métodos tecnológicos de punta se pueden digitalizar a través de programas de computación y tener sistemas de información georreferenciados.

Para el caso de encuestas, se utiliza la base de datos del censo como marco muestral y a partir de métodos estadísticos se seleccionan muestras, las cuales también deben de ser actualizadas en campo.

1.3.4 Cargas de trabajo o áreas de empadronamiento (AE)

Dependiendo del tiempo estipulado para la entrevista en campo se deben de realizar las cargas de trabajo. Como insumo para esto se utilizan las actualizaciones cartográficas que determinaron el número de viviendas por lugar poblado. Con base en lo que se desea trabajar por empadronador o encuestador, se hacen áreas de trabajo con el tamaño de viviendas deseado, las cuales forman unidades de medida más pequeñas para supervisar.

Este trabajo se puede realizar sobre el archivo base actualizado (mapas, planos y croquis) o si se cuenta con programas especializados que trabajen sobre *software* de sistemas georreferenciados se pueden automáticamente hacer las cargas de trabajo.

1.3.5 Reproducción de material

Una vez definidas las cargas de trabajo, se tendrá que reproducir, para entregarle a cada supervisor y empadronador, una representación gráfica (mapa, plano o croquis) del área que tendrá a su cargo. Esto se debe de empacar apropiadamente y enviar al campo para su distribución según la pirámide de trabajo de campo.

1.3.6 Empadronamiento

Conociendo las cargas de trabajo y el área donde se elaborara el levantamiento de datos, se debe empezar a recolectar la información. Para el empadronamiento, es necesario que se tenga buena capacitación y supervisión con el fin de que el personal investigue las variables tal y como fueron planificadas.

En el caso de los censos, por su magnitud, no da tiempo de realizar correcciones en campo, por lo que se tendrán que hacer a través de programas especializados. Por su parte en las encuestas (principalmente las que son de varios días en una misma unidad de estudio) se pueden aclarar algunas respuestas con el propio informante en el campo y dejar sólo las consistencias básicas para que se efectúen en las oficinas centrales.

2 TEORÍA DEL PROCESAMIENTO DE DATOS

2.1 Procesos implicados en el procesamiento

2.1.1 Consideraciones básicas

El procesamiento de datos es el conjunto de operaciones que traduce la información recolectada en un censo en un conjunto útil y oportuno de informes estadísticos. Un censo nacional de población y habitación es, casi invariablemente, la operación de procesamiento de datos más cara y extensa que puede emprender un país. Es más, los períodos entre censos, normalmente de 10 años, son tan largos que la transferencia de conocimiento y técnicas de un censo a otro es mínima, particularmente en vista de los cambios tecnológicos que tienen lugar en el intervalo. En efecto, el procesamiento de cada censo nacional debe tratarse como una operación extremadamente grande y como si se realizara por primera vez. Sin embargo, todos los procesos implicados en el procesamiento siempre son los mismos por lo que se deben de conocer a cabalidad.

El secreto para que tenga éxito una operación de procesamiento de datos es mantenerla continua y directa y lo más sencilla posible. El plan total para el procesamiento debe determinarse con bastante anticipación al período de empadronamiento.

El plan debe incluir: (a) un calendario basado en los niveles de producción esperada y volumen de trabajo; (b) la cantidad de procesamiento manual requerido, basado en el diseño del cuestionario y en los procedimientos de empadronamiento; (c) el tipo de equipo que se va a usar, incluyendo sistemas de conversión de datos, equipo de computación (“*hardware*”) y paquetes de programación (“*software*”); (d) tipo y cantidad de materiales tales como CDs para el computador; (e) requerimientos de personal, incluyendo procedimientos de contratación y de capacitación; (f) cantidad y forma de descentralización del procesamiento; (g) organización y flujo de trabajo y (h) procedimientos de control, incluyendo control de operaciones, control de calidad y planes de contingencia para incidentes pequeños y grandes.

2.1.1.1 Elementos principales del procesamiento de datos

El procesamiento de datos en la mayoría de los censos requerirá completar con éxito las operaciones descritas a continuación. Con excepción de la primera operación, que se completa poco después del empadronamiento de terreno, todas las otras operaciones se traslapan en el tiempo. Cuando se empieza el procesamiento censal, se puede esperar que todas las operaciones estén realizándose simultáneamente. Las operaciones principales incluyen:

- 1) Recibir y controlar los cuestionarios completados para asegurarse que se han recibido los cuestionarios correspondientes a cada una de las cargas de trabajo o áreas de empadronamiento.
- 2) Crítica y codificación manual de los cuestionarios, incluyendo recuentos de población y de cuestionarios enumeradas, por AE, con propósitos de control.
- 3) Conversión de los datos de los cuestionarios a formas legibles para su procesamiento (codificación).

- 4) Ingreso de los datos al computador.
- 5) Crítica automática (por computador) de datos.
- 6) Tabulación de los datos y producción de cuadros para publicación.

Las operaciones anteriores, de 1 a 3, involucran procesos de oficina o manuales a gran escala. Las operaciones 4 a 6 se consideran procesos de computador o relacionados con el computador.

Actualmente el punto 3 también se puede considerar como proceso relacionado con el computador y se explicará más adelante.

2.1.1.2 Procesamiento manual versus procesamiento por computador

Los procedimientos más efectivos de procesamiento son aquéllos que minimizan las operaciones de procesamiento manual involucradas en criticar, codificar y tabular los datos. Se ha demostrado que mientras más procesamiento manual se hace, más demora la publicación de los resultados censales y mayores son las inconsistencias debidas a error humano.

En la mayoría de los casos, las operaciones de crítica y tabulación se pueden hacer por computador; éste puede realizar estas operaciones con más eficiencia y precisión. Un buen diseño del cuestionario puede minimizar la cantidad de codificación manual (si se utiliza el método de codificación previa).

Confiar en el procesamiento por computador en el grado indicado requiere que el país cuente con personal capacitado en las técnicas actuales de programación. Además, se requiere un sistema de computador de tamaño y capacidad suficientes para ejecutar la metodología de procesamiento moderna.

El énfasis en el uso de los procesos por computador, sin embargo, no pretende sugerir que las operaciones de tabulación manual, además de las operaciones manuales requeridas para obtener resultados preliminares de población, no deban considerarse. Por ejemplo, si es importante para la determinación de la política nacional obtener información sobre la distribución de la población por edad y sexo tan rápidamente como sea posible, se pueden desarrollar técnicas para hacer eso e incorporarse dentro de la operación del recuento manual empezando desde los empadronadores hasta que la estructura organizacional lo permita.

Un punto importante para recordar es que hay costos adicionales asociados con ciertas operaciones manuales que podrían hacerse a máquina como parte de otro procesamiento mecánico. Será necesario emplear más personal de oficina; se requerirá procedimientos adicionales de verificación y de control de calidad y los resultados diferirán, virtualmente para cada área geográfica, de los producidos después en las tabulaciones por computador.

2.1.1.3 Calendario de actividades

Generalmente se subestima el lapso de tiempo necesario para procesar un censo o una encuesta, debido principalmente, a que no se reconoce del todo la complejidad de las operaciones de procesamiento y a que no se aprecia del todo la importancia de elaborar un calendario de actividades. Debe elaborarse un calendario detallado para el procesamiento de datos e incorporarse al calendario general de todas las actividades censales.

2.1.2 Crítica

Toda persona que esté familiarizada con censos y encuestas sabe que es imposible obtener datos completos y precisos para todos los *items* en todos los cuestionarios. En la mayoría de los países, parece haber poca falsificación deliberada de respuestas ya sea por parte de informantes o de empadronadores. Sin embargo, en algunas respuestas ocurren sesgos sistemáticos, probablemente como resultado de diferencias humanas y factores culturales. También se introducen errores, inadvertidamente, durante las operaciones de codificación manual y conversión de datos.

Por crítica se entiende la inspección y corrección de las respuestas de acuerdo con normas predeterminadas. La crítica se puede hacer como parte de la operación manual de crítica y codificación de la oficina central o puede hacerse en otras etapas del censo.

Por ejemplo, la crítica usualmente se realiza durante la recolección de datos, durante el procesamiento mecánico y durante la revisión técnica anterior a la publicación de los resultados. Si los resultados censales no se critican, es probable que contenga cierta cantidad de datos sin sentido.

2.1.2.1 Tipos de crítica

Como se hizo notar previamente, la crítica puede hacerse, y frecuentemente es así, en casi todas las fases de un censo o encuesta. A continuación se enumeran los tipos principales de crítica.

2.1.2.1.1 Crítica por el empadronador

La crítica por el empadronador usualmente consiste en una revisión de las anotaciones hechas en cada cuestionario al concluir la entrevista.

2.1.2.1.2 Crítica de terreno

En algunos casos, el supervisor de terreno y el personal de la oficina de terreno pueden también efectuar una crítica antes de enviar los cuestionarios a la oficina central. Pueden ser cuestionarios aleatorios.

2.1.2.1.3 Crítica manual por la oficina central

Otro tipo de crítica es la efectuada por la oficina central. Esta crítica sigue instrucciones precisas y detalladas. Difiere, por ejemplo, básicamente de la crítica por empadronador; en esta última, las correcciones se hacen mientras se está en contacto con el informante.

2.1.2.1.4 Crítica por computador

La crítica por computador, como la manual, sigue un conjunto preciso de reglas. En la crítica por computador, la corrección de una anotación puede controlarse examinando las condiciones casi simultáneamente.

Además, la imputación de características que faltan pueden basarse en técnicas de asignación tales como la “asignación dinámica” (“*hot decking*”), que proporcionará resultados más consistentes.

Sea que un empleado o un computador esté haciendo las correcciones, es esencial mantener un registro del número y tipo de correcciones hechas a cada *item* para evitar la posibilidad de usar datos que sean en gran parte fabricados. Este registro se llama, usualmente, diario de crítica. Puede ser producido automáticamente durante la crítica por computador.

2.1.2.2 Tipos de errores

Los tipos de errores que pueden aparecer en los cuestionarios individuales caen dentro de varias categorías. En algunos casos, un error puede clasificarse en más de una categoría.

2.1.2.2.1 Omisiones

Estas son, sencillamente, descuidos por parte de los empadronadores al insertar anotaciones. En muchos casos, la respuesta correcta se puede determinar por medio de otras anotaciones en el cuestionario de la misma persona o unidad de vivienda.

2.1.2.2.2 Anotaciones imposibles

Son anotaciones imposibles aquéllas que son ilógicas o prohibidas por el sistema de codificación. Por ejemplo, si el código 1 representa “hombre” y el código 2 “mujer”, cualquier otro dígito constituye una anotación imposible. La asignación de un código imposible ocurre con mucha frecuencia en la operación de conversión de datos.

2.1.2.2.3 Anotaciones inconsistentes

Ocurren las anotaciones inconsistentes cuando dos o más *items* que deben tener entre sí una relación determinada, no la tienen.

2.1.2.2.4 Magnitudes irracionales

Las anotaciones de este tipo pueden no ser necesariamente errores, pero no parecen razonables. Por ejemplo, una anotación para un trabajador no especializado puede presentar un ingreso excesivo para el promedio de ese determinado grupo ocupacional. Una anotación así puede estar correcta, pero merece investigación.

2.1.2.3 Principios para elaborar programas de crítica

El propósito de la crítica es hacer los datos tan representativos del universo como sea posible, eliminando omisiones, anotaciones imposibles y anotaciones inconsistentes. A continuación hay algunos principios fundamentales que deben seguirse:

- (1) Hacer un mínimo de cambios en los datos registrados originalmente.
- (2) Eliminar inconsistencias obvias en las anotaciones.
- (3) Hacer anotaciones para ciertos *items* que falten, usando otras anotaciones para la unidad de vivienda u otras personas del hogar como guía, siempre de acuerdo con procedimientos específicos.

Para corregir existen dos métodos de corrección:

2.1.2.3.1 Método de “asignación estática” (“*cold deck*”)

Este método utiliza imputaciones de respuestas válidas conocidas previamente (por censos o encuestas previas) y se utilizan porcentajes de aplicación para mantener la estructura de la información conocida.

2.1.2.3.2 Método de “asignación dinámica” (“hot deck”)

Otro método de imputación es el uso de la respuesta válida para la última persona en el archivo con características similares. Una “asignación dinámica” inicialmente empieza como “asignación estática”. Se elabora un conjunto de respuestas lógicas para usarse en el primer registro procesado que contiene anotaciones no válidas. La asignación estática es preparada por especialistas en temas que usan datos de censos anteriores, encuestas y otro material relevante.

2.1.3 Codificación

La codificación es en gran parte un proceso de traducir clasificaciones de palabras a números de modo que los datos puedan tabularse a máquina. La codificación puede combinarse con la operación de crítica o realizarse como proceso separado. Que la crítica y la codificación se combinen o sean dos operaciones distintas depende de la naturaleza del cuestionario y de la complejidad de clasificación de las anotaciones.

2.1.3.1 Precodificación y auto codificación

Si fuese posible, es ventajoso diseñar los cuestionarios usando *items* precodificados o auto codificados.

Es más eficiente que el empadronador encierre en círculo un código numérico para una respuesta, que anotar una respuesta y codificar la entrada en la oficina. Puede que el cuestionario no tenga espacio suficiente para que el empadronador anote toda la información requerida como para codificar una respuesta escrita. Además, las respuestas que se anotan escritas requieren más procesamiento manual, lo cual es más caro y está sujeto a más errores.

2.1.3.2 Codificación por computador

Actualmente se han desarrollado sistemas expertos que utilizan el reconocimiento de lenguaje para codificar y clasificar las diferentes preguntas abiertas que hay en un cuestionario.

2.1.3.3 Recodificación

La recodificación es simplemente la combinación de códigos para varios *items* relacionados en un solo código para conveniencia y eficiencia en la tabulación. En algunos casos, es imposible formular una sola pregunta para recolectar la información necesaria precisa para categorizar a las personas para la tabulación; de hecho, puede ser necesaria más de una pregunta. Por ejemplo, una sola pregunta no identifica, usualmente, a una persona ocupada, desocupada o que no es económicamente activa. Debido a que una cantidad de tabulaciones usan estas categorías, es conveniente examinar varias preguntas relevantes y asignar al individuo un recódigo (o código sumario). Generalmente, la recodificación se realiza con mayor exactitud y eficiencia por medio del computador que manualmente.

2.1.4 Métodos de conversión del dato

Una de las operaciones más caras y demoradas es la de convertir datos del cuestionario a una forma adecuada para procesamiento por computador.

El método y equipo afectará a muchos otros elementos del programa censal, como diseño del cuestionario, procedimiento de empadronamiento, tipo de personal requerido y presupuesto. Por lo tanto, la decisión sobre el método y equipo que se va a usar para esta operación debe tomarse al principio de las etapas de planificación del censo.

2.1.4.1 Examen de la experiencia previa

Los planificadores deben examinar la historia de procedimientos del último censo por si hubiera surgido algún problema con el método o equipo usado. Si el método usado tuvo como resultado problemas con la calidad de los datos o atrasos en el procesamiento, los planificadores deben investigar otros métodos de conversión de datos. Aunque el método usado haya sido adecuado, debe evaluarse el equipo que se tiene a mano. Si es obsoleto o está en malas condiciones de funcionamiento, puede ser necesario adquirir equipo diferente o más moderno. Por otra parte, si el equipo funciona bien, no es necesario tomar ninguna acción; o tal vez lo único necesario será adquirir equipo adicional del mismo tipo. Antes de tomar una decisión definitiva, se debe hacer contacto con el fabricante para obtener información sobre la disponibilidad de piezas y equipo de reemplazo.

2.1.4.2 Evaluación de los métodos

Hay literalmente cientos de sistemas para convertir los datos a forma utilizable para procesamiento por computador.

Estos sistemas caen dentro de dos categorías principales: (a) aquéllos que requieren alguna forma de registro por teclado y (b) aquéllos de lectura óptica. Conviene limitar, lo más rápidamente posible, el campo de los métodos que se van a considerar.

Como consideraciones para eliminar se debe tomar en cuenta la existencia de apoyo técnico tanto en *hardware* como *software*, tipo de personal disponible, capacidades de espacio, etc.

2.1.4.3 Estimaciones de personal y equipo

Una vez tomada la decisión sobre el tipo de información, debe calcularse la cantidad requerida de recursos y operadores, usando el volumen esperado de trabajo, tasas de producción y tiempo programado para la conversión de datos.

Las estimaciones deben incluir la cantidad de personal para supervisión y control y la cantidad de personas y recursos necesarios para verificación.

La siguiente decisión que debe tomarse es respecto a operar o no el equipo en dos turnos, reduciendo así la cantidad de aparatos requeridos a más o menos la mitad. Hay una cantidad de factores que determinan esta decisión.

- 1) Hay diferencia entre usar X aparatos durante Y meses versus usar $0.5 X$ aparatos en dos turnos durante Y meses. El arriendo de las máquinas debe ser menor debido a que son menos y se reduce la cantidad de espacio.
- 2) El fabricante debe ser capaz de proporcionar servicio con base en dos turnos para que este plan sea factible.
- 3) El personal debe estar disponible para trabajar varios turnos. Es costumbre pagar una tasa ligeramente más alta por trabajo nocturno, lo que neutralizaría parte de los ahorros en el costo de los aparatos.

2.1.5 Programas de computador

Para diseñar, elaborar y probar programas de crítica de datos por computador, se espera que un equipo de programadores altamente calificados trabaje por varios años, incluyendo por lo menos un año después de recibir especificaciones más o menos completas. Los programas de tabulación adecuados necesitan la misma cantidad de tiempo y de personal.

La mayoría de los países no tienen ni tiempo ni personal para una empresa como esa. Por lo tanto, a principios de la etapa de planificación debe investigarse la disponibilidad de paquetes de programas generales para criticar y tabular el censo. Aunque haya suficientes programadores capacitados disponibles, probablemente su tiempo se puede emplear mejor en elaborar programas especiales para los que no haya programas generales.

2.1.5.1 Selección de paquetes de programas

El primer paso para determinar si un paquete es adecuado es saber de él todo lo que sea posible. La mayoría de los que elaboran paquetes tienen programas de instrucción para iniciar a los usuarios potenciales en las posibilidades y restricciones de su producto específico.

La asistencia a una sesión de instrucción es la mejor manera de aprender lo suficiente como para hacer una evaluación inteligente del paquete. Cada programa disponible debe ser evaluado dando respuestas a las siguientes preguntas:

- (1) ¿Se desarrolló el programa por personal con experiencia en censos o encuestas?

- (2) ¿Cuál es la capacidad o voluntad del creador del programa para ofrecer instrucción y ayuda de instalación? ¿Con cuánta prontitud puede estar operando el paquete?

- (3) ¿Hasta qué punto está completo el paquete y cuánta programación auxiliar se necesita?
- (4) ¿Cuál es la experiencia de otros organismos estadísticos o de otros países en el uso del paquete para censos o encuestas?
- (5) ¿Qué tipo de personal se requiere para usar el paquete?
- (6) ¿Cuánta preparación y de qué tipo se necesita cada vez que se vuelva a usar? ¿Hasta qué punto está el programa libre de error?
- (7) ¿Qué tamaño y condiciones debe tener el computador para que el paquete funcione efectivamente?
- (8) ¿Qué cantidad y tipo de equipo adicional debe adquirirse para usar el programa?
- (9) ¿Cuál es la tasa de producción? ¿Cuánto tiempo de funcionamiento y de preparación del computador se requiere?
- (10) ¿Hasta qué punto está completa la documentación para el paquete? ¿Se describen totalmente sus características en un manual bien escrito para los usuarios?

Muchas firmas programadoras comerciales venden paquetes para procesar datos. Desafortunadamente, pocas de estas firmas tienen especialistas en computación experimentados que estén familiarizados con las características y requisitos para criticar y tabular los datos censales.

Hay paquetes disponibles a través de las Naciones Unidas y de organizaciones de países con años de experiencia en procesamiento por computador de datos censales. Se debe investigar diversas fuentes y evaluar sus programas. El conocimiento de las posibilidades y limitaciones de diversos paquetes de programas es de poco valor si no se conocen los requisitos de procesamiento del censo. No son necesarias las especificaciones de procesamiento definitivas, pero es esencial conocer los requisitos generales para determinar si el paquete de programas satisface las necesidades censales.

2.1.5.2 Prueba de programa de computador

Antes de tomar una decisión definitiva sobre el paquete de programas, se debe planificar una prueba con una muestra de datos o con los datos de la prueba censal de preferencia en el sistema de computador que se va a usar para el procesamiento censal. Para hacer esto, las especificaciones para la crítica y tabulación deben ser lo más completas posibles.

Un posible método para hacer esta prueba sería preparar un conjunto de cuestionarios que tuvieran una variedad de respuestas, con algunas omisiones, inconsistencias y anotaciones irracionales. El objetivo sería asegurar que el programa opere correctamente.

2.1.5.3 Elaboración de programa de computador

Aunque la organización de estadística puede realizar economías considerables por medio de la instalación y uso de programas de computador generalizados en paquete, siempre será necesario elaborar algunos programas de computador especializados para las necesidades particulares del censo. Tal vez el equipo de registro de datos o las especificaciones para la tabulación requerirán que los datos vuelvan a ser puestos en formato antes de ser procesados.

Tal vez el programa en paquetes no ejecute ciertos cálculos o no produzca ciertas tabulaciones que se consideran muy importantes. Debido a que el programa en paquete es general, nunca se ajustará a todas las necesidades especializadas de un censo. En consecuencia, la institución de estadística debe planificar, elaborar, escribir y probar los programas adicionales o auxiliares que se necesiten.

2.2 Administración del procesamiento

2.2.1 Prueba de sistemas

A medida que se desarrollan los procedimientos para el procesamiento de datos, aquéllos deben probarse para asegurar que son comprensibles y que su aplicación produce los resultados deseados.

Se deben hacer pruebas de las operaciones de crítica y codificación manuales, los procedimientos de control de operaciones, los procedimientos de control de calidad, la operación de registro por teclado, los programas de computador, las operaciones de revisión del diario y los procedimientos para revisar los cuadros de publicación.

Como se mencionó antes, los datos de las pruebas y ensayos de terreno pueden usarse para probar el sistema o se puede preparar una “muestra de indicios”. En una muestra de indicios, se desarrollan los datos de la muestra creando varios cientos de hogares hipotéticos y preparando cuestionarios censales para ellos.

Se debe tener cuidado de crear hogares de todo tipo, de modo que se puedan probar los procedimientos tanto para los casos ordinarios como para los extraordinarios. Se debe incorporar al fichero de la prueba información inconsistente y omisiones.

Las operaciones que no avancen fácilmente pueden alterar completamente el calendario, causar déficit en el presupuesto y afectar en forma adversa la calidad de los datos. Debe probarse, evaluarse, revisarse y volverse a probar cada paso en las operaciones de procesamiento de datos; en seguida se debe hacer una prueba de todas las operaciones como un todo. Para la prueba final, es conveniente el uso de los cuestionarios de un censo o encuesta experimental. Esperar hasta que haya datos “vivos” disponibles del campo mismo será demasiado tarde; usualmente, las revisiones son demasiado complejas para hacerse rápidamente.

Una parte integral de las pruebas debe ser la manutención de los registros de producción y de control de calidad para las principales operaciones de oficina y de registro por teclado. Sin esa información, será difícil estimar el volumen de personal, el espacio y el equipo de trabajo requeridos.

Igualmente importante, los resultados de la prueba identificarán esas etapas de las operaciones que sean excesivamente difíciles y demoradas; de este modo, las pruebas pueden llevar al desarrollo de métodos mejorados y alternativos.

2.2.1.1 Operaciones de revisión, crítica y codificación manuales

En las primeras etapas de prueba de las operaciones de revisión, crítica y codificación, el personal profesional debe probar los procedimientos, analizar las dificultades y revisar lo que sea necesario. En las etapas posteriores y finales, las pruebas deben ser efectuadas por personas del mismo tipo y nivel que las que realicen las operaciones mismas del censo.

Los empleados deben ensayar los procedimientos de recibo y control de ingreso; los revisores y codificadores deben probar los procedimientos de crítica y codificación; los mensajeros y empleados de control deben someter a prueba los procedimientos de envío y de control, etc. Es de la mayor conveniencia que los supervisores potenciales se desempeñen primero como empleados, revisores y codificadores, puesto que esto les permitirá conocer las necesidades de fuerza de trabajo para procesamiento, sugerir mejoras en los procedimientos y manejar en forma más efectiva las operaciones por las cuales tienen principal responsabilidad.

2.2.1.2 Operaciones de registro de datos y computación

Durante la selección de los aparatos para registro de datos, paquetes de programas y equipo de computación, deben realizarse una serie de pruebas de diversos equipos y programas. Estas pruebas deben efectuarlas por las personas responsables de tomar las decisiones, asistidas por proveedores o sus representantes.

Pero algún tiempo antes del procesamiento censal mismo, se debe probar, en una situación lo más parecida posible a la esperada en el censo, el procedimiento de registro de datos, el programa de computación y el computador. Debe probarse el registro de datos, la consistencia y la tabulación.

El producto de la prueba debe ser impreso y sujeto a revisión por personal que estará efectuando esta tarea durante el censo. Esto proporcionará una medida del tiempo que se requerirá para esta función y probará también la calidad de la impresión.

2.2.2 Requisitos y funciones del personal

En la mayoría de los países, el personal profesional, que incluye a los especialistas en procesamiento de datos, debe aumentarse considerablemente para emprender el censo. Esto puede lograrse por medio de transferencias desde otros organismos de gobierno, reclutando trabajadores temporales con base en períodos cortos o contratando personal independiente o de firmas procesadoras de datos.

2.2.2.1 Personal profesional y técnico

Este personal debe incluir analistas de sistemas y programadores. Si fuera posible, debe haber un conjunto de analistas y programadores con alguna experiencia en censos o encuestas para planificar, diseñar o describir los procedimientos y supervisar. Se puede contratar personal adicional y capacitarlo mientras trabaja. El volumen de personal que se necesite depende en parte del grado en que se usen paquetes de programas generalizados y de la cantidad de programación auxiliar necesaria.

2.2.2.2 Personal de registro o revisión por teclado

En general, se aplican los mismos principios a la selección del personal y supervisores de registro por teclado, incluyendo la notificación de que se aplicarán normas de control de producción y de calidad. El personal de registro por teclado debe tener destreza manual; usualmente las personas hábiles en mecanografía se desempeñan bien.

2.2.3 Sistemas de control

Además del calendario de actividades, que sirve de base para la medición total del avance, se necesitan controles efectivos de las operaciones día a día. Se deben establecer sistemas de control de operaciones de calidad. El éxito de las operaciones de procesamiento a gran escala depende tanto de lo adecuado de estos sistemas de control como del desarrollo de procedimientos efectivos de procesamiento.

2.2.3.1 Control de operaciones

Incluso un plan de procesamiento bien diseñado puede dejar de alcanzar los objetivos de oportunidad y precisión, si no se sigue con vigor y eficiencia. Aunque no hay sustituto para una buena administración, hay una cantidad de pasos para hacer más efectivo el proceso administrativo y contribuir a asegurar un avance apropiado y continuo.

2.2.3.1.1 Normas de producción y calendario

Las normas de producción son la clave para cumplir el calendario de procesamiento y completar la operación dentro de los costos presupuestados.

Debe determinarse cuánto de cada operación tiene que completarse dentro de un período de tiempo (día, semana, mes) para cumplir las fechas que se establecieron en el calendario de actividades.

Para las operaciones de oficina y de registro por teclado, se pueden establecer tasas de producción significativas sólo con base en experiencia con tipos similares de cuestionarios con una cantidad equivalente de procesamiento o por medición directa. La última utilizaría personal capacitado para criticar, codificar y registrar por teclado un conjunto de cuestionarios de prueba.

El calendario de producción ayudará en la preparación de asignaciones semanales o mensuales de presupuesto detalladas. También proporciona referencias con las cuales se puede medir el avance.

2.2.3.1.2 Diagramas de flujo

El procesamiento de un censo nacional o de encuestas, cuando se considera en todos sus aspectos, es una operación demasiado compleja para que un administrador recuerde todos los detalles.

Para asegurar un proceso ordenado y controlable, no sólo deben señalarse los detalles de cada operación en forma precisa, sino que se debe mostrar el flujo de trabajo y las relaciones entre las operaciones.

Los diagramas de flujo de trabajo, que presentan el rango total de las actividades del procesamiento de datos censales en forma visual, describen los requisitos de operaciones de una manera que los detalles específicos y las interrelaciones están claras para todos.

Los diagramas de flujo sacarán, también, rápidamente a luz elementos de operaciones de indebida complejidad que pueden entonces reexaminarse con el objetivo de desarrollar procedimientos más sencillos.

Los diagramas de flujo son una herramienta importante para dirigir el procesamiento. La elaboración, desde un principio, de buenos diagramas que cubran tanto las actividades de operaciones principales como los detalles de procesos específicos, es un ingrediente esencial de un sistema de procesamiento de datos eficiente y efectivo.

2.2.3.2 Control de calidad

Un objetivo principal de todo censo y encuesta es producir estadísticas confiables. Se usan procedimientos de control de calidad para medir la calidad del rendimiento y para limitar el número de errores a un grado aceptable.

El establecimiento de procedimientos de control de calidad para las operaciones de procesamiento de datos es extremadamente importante, especialmente para las operaciones de procesamiento y registro por teclado manual. La crítica y codificación manuales están sujetas a un alto grado de sesgo y error humanos.

Las normas y procedimientos de control de calidad varían de acuerdo con las disposiciones sobre el nivel de precisión exigido por los productores y usuarios de las estadísticas.

En todos los casos, el objetivo no es garantizar un censo libre de errores (lo cual es imposible), sino asegurar que esos errores no afectarán la validez estadística de los resultados censales.

2.2.4 Ambiente físico

Un ambiente físico adecuado de trabajo es condición necesaria para una operación de procesamiento de datos eficiente. Los equipos de computación y de registro de datos requieren de un ambiente con temperatura y humedad controlada. Los escaners, en particular, se pueden atrancar o quedar inutilizados si están expuestos a frecuentes y excesivos cambios de humedad.

También debe considerarse el ambiente físico para el personal de oficina y de digitación.

Las tasas de producción previstas y los niveles de calidad serán difíciles de alcanzar si el ambiente físico es insatisfactorio. Se debe disponer de suficiente espacio para hacer posible un flujo de trabajo parejo y seguro, incluyendo espacio para el almacenamiento temporal de cuestionarios adyacentes a las operaciones de oficina y digitación. Una iluminación y ventilación adecuadas son esenciales para un trabajo de buena calidad.

Si es difícil obtener un espacio adecuado, se debe considerar la posibilidad de establecer las operaciones en dos turnos. Si se debe controlar el ambiente, es económico hacer uso intensivo del espacio con operaciones en dos turnos.

2.2.5 Documentación del sistema

La documentación escrita de todas las decisiones operativas, planes y procedimientos es absolutamente necesaria.

Preferentemente, esto debe hacerse a medida que se toman las decisiones y que evolucionan los planes, puesto que será virtualmente imposible hacerla en una fecha posterior basándose en la memoria y en notas. De otro modo, no habrá explicación adecuada en lo que se refiere a por qué se tomaron ciertas decisiones o se escogieron ciertos procedimientos. Durante la etapa de planificación se deben emitir indicaciones formales para asegurar la documentación de todas las operaciones.

Debido a la interrelación de las operaciones de procesamiento, se debe mantener la documentación al día y distribuirse a las personas que corresponda. Esto se aplica especialmente a las modificaciones en los procedimientos y a los cambios en las especificaciones técnicas. Con seguridad habrá confusión y atrasos costosos si no se prepara una documentación adecuada y oportuna.

La documentación a la que se hace mención anteriormente servirá como base para preparar el capítulo del procesamiento de datos en la historia de procedimientos. La documentación debe incluir:

- 1) Decisiones básicas sobre procesamiento de datos, incluyendo obtención de equipos y adquisición de programas en paquete.
- 2) Procedimientos de recibo, ingreso y control de operaciones.
- 3) Procedimientos de revisión del diario, crítica y codificación manuales.
- 4) Procedimientos de registro por teclado.
- 5) Procedimientos de procesamiento por computador de crítica, codificación y tabulación.
- 6) Procedimientos de control de calidad.
- 7) Funciones y requisitos del personal, incluyendo actividades de capacitación.
- 8) Presupuesto, costos y persona-hora gastadas por cada operación.

La historia de procedimientos sirve como registro de logros y como una excelente base para planificar el censo o encuesta próximos.

3 SOFTWARE PARA PROCESAMIENTO DE DATOS

3.1 Características del *software*

Para poder trasladar la información proveniente del campo a información manipulable en bases de datos existe el método tradicional por teclado. Este método requiere de personal adiestrado para ingreso de datos y de programas especializados.

3.1.1 Programas especializados

Es importante notar que se pueden desarrollar programas de captura de datos masiva en cualquier lenguaje de programación; sin embargo, para que se tenga éxito en el proceso de datos (datos oportunos y de calidad) es necesario que se utilicen programas especializados que tengan como mínimo las siguientes características:

- a) Definición de diccionarios basados en la boleta de campo.
- b) Herramienta para desarrollo de pantallas de captura basadas en los diccionarios.
- c) Programas para consistencia de datos basados en diccionarios.
- d) Programas para tabulaciones simples basados en diccionarios.

Con programas desarrollados ad-hoc es poco probable –debido al tiempo y esfuerzo que conlleva- que se logren unificar las cuatro características que debe cumplir un programa especializado para el procesamiento de datos censales y encuestas.

Además puede ser que se haga un buen desarrollo pero que con el paso del tiempo y los diferentes requerimientos de investigación quede obsoleto; por lo tanto, es importante que se investiguen herramientas que se utilizan en los diferentes países, elaboradas y mejoradas por el paso del tiempo por instituciones que regulan y normalizan la ejecución de censos y encuestas.

Debido al avance tecnológico, actualmente se puede suplir la digitación con el uso de escáneres que interpretaran los datos. Existen en el mercado muchos programas que utilizan motores de reconocimiento óptico que se están utilizando actualmente para procesar censos y encuestas. Estos programas también deben cumplir con las características expuestas anteriormente, sólo que en lo relacionado con el reconocimiento óptico. Los sistemas de reconocimiento óptico normalmente deben incluir los siguientes procesos:

Proceso de recepción y preparación de material: se reciben los documentos agrupados por donde fueron tomados en campo, se valida que toda la documentación ingrese a la unidad de procesamiento. Después cada lote de cuestionarios se prepara de forma que los documentos puedan ser escaneados y se valida la consistencia del número de formularios escaneados versus lo que se tenía contemplado que debía entrar.

Sistema de escaneado: se escanean los documentos; simultáneamente se verifica la calidad de las imágenes y se valida que el número de imágenes corresponda al número de documentos del lote recibidos.

Sistema de reconocimiento: se envían las imágenes del lote a la cola de reconocimiento. Los lotes son tomados en modo FIFO por medio del motor de reconocimiento de sus datos. Los datos y la información operacional son registrados en el *software*.

Sistema de verificación: se refiere a los datos de lotes cuya totalidad de caracteres fueron interpretados con un nivel de confianza mayor a 90%. Los lotes que contienen caracteres cuestionados son alimentados a la cola de verificación, de ese último, o el verificador tomará los lotes y procederá a verificar. La verificación consiste en confirmar, corregir o registrar la información de cada campo con caracteres cuestionados contra el campo imagen correspondiente.

Sistema de transcripción: se transcriben los lotes que contengan más del 30% de boletas deterioradas y no pueden ser tratadas por captura óptica. Se digitarán y verificarán en modo tradicional.

Sistema de exportación: para poder dar consistencia a los datos puede ser necesario exportarlos a diferentes formatos, por lo que es necesario que se tenga la opción de exportación.

3.2 Software por teclado

3.2.1 IMPS

El *Integrated Microcomputer Processing System* (IMPS) es uno de los sistemas de mayor desempeño en procesamiento de datos de censos y encuestas. Cuenta con definición de diccionarios, entrada de datos, consistencia, tabulación y otros componentes secundarios como distribución del dato, estadísticas y análisis del proceso de captura, teniendo los módulos independencia para utilizarse por separado.

IMPS tuvo sus comienzos desde los años 80 con desarrollo y esfuerzos independientes realizados por entidades como CELADE y de instituciones estadísticas gubernamentales; sin embargo, el Buró de Censos de los Estados Unidos compró los derechos y realizó la primera versión integrada.

En 1997 realizaron una revisión y desarrollaron la primera versión para Windows, y se mejoró a principios de 2000. Este es un programa de distribución gratuita.

Este es un *software* monousuario por lo que hay que realizar procesos de integración de datos en red; las características técnicas de las computadoras para que se ejecute son mínimas:

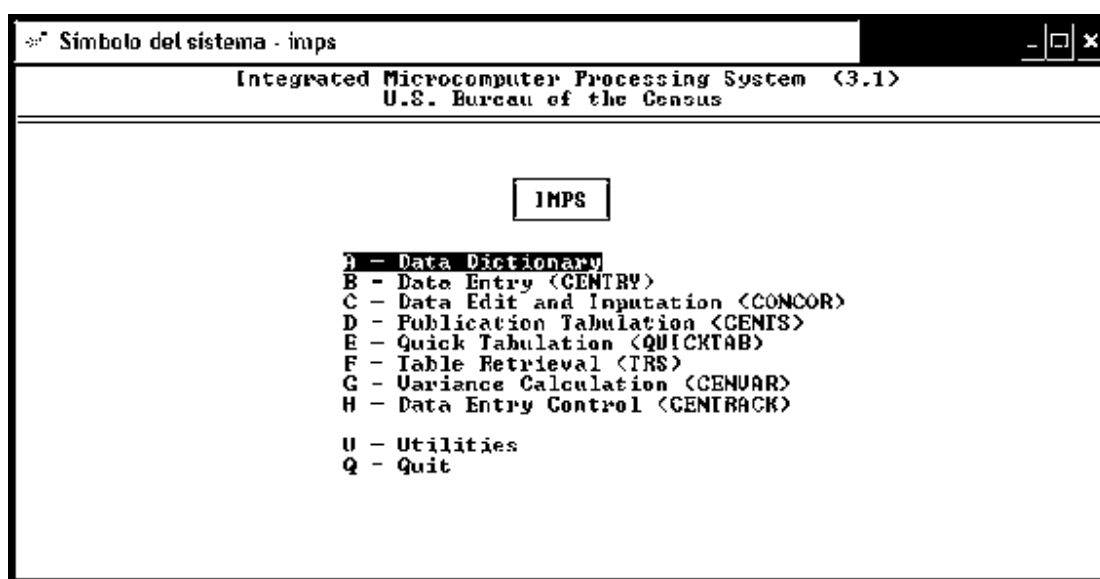
- Procesador pentium
- 16 MB de RAM
- SVGA monitor (PCI o VESA)
- *Mouse*
- 16 MB de espacio libre (5 MB si IMPS 3.1 está ya instalado)
- Microsoft Windows 95 o superior

3.2.1.1 Componentes principales de IMPS

Menú principal de IMPS

Al ingresar a IMPS, se muestra la siguiente pantalla:

Figura 1. Menú principal de IMPS



A continuación se describen los componentes de mayor importancia en IMPS:

Data Dictionary (Diccionario de datos)

Este Módulo se utiliza para definir las características del archivo de datos que va a ser procesado. Esta definición, que es trasladada a computadora, se utiliza en el Módulo de entrada de datos, consistencia y tabulación, por lo que antes de utilizar cualquier herramienta dentro de IMPS es necesario definir cómo se van a trabajar los datos.

Es aquí donde se definen las diferentes variables que se utilizan dentro de la boleta en campo.

Normalmente se tiene un identificador de boleta (que incluye el lugar donde se toman los datos) y luego vienen todas las variables por investigar agrupadas por tema. Siguiendo esta característica de definición, para declarar un diccionario es necesario definir lo siguiente:

- Definir los campos comunes a toda la boleta (o de identificación): cuando se hayan definido los campos comunes se debe ingresar al menú “**Common**” para definir las diferentes variables que identifican a una boleta.
- Definir los registros con sus variables: un registro es un conjunto de variables que identifican un tema que se puede agrupar. Cuando ya se han definido los diferentes registros, se debe de ingresar al submenú “**Records**” donde se coloca un nombre con el cual se hace referencia al registro en los demás módulos de IMPS; se debe indicar si es requerido el registro con sus variables (es decir que es necesario para completar la boleta) y cuántos registros de ese tipo máximo se aceptan como válidos. Una vez definido el registro, se deben indicar todas las variables que lo componen.

Para definir un diccionario en IMPS es necesario conocer los siguientes términos:

Data Ítem: contiene la respuesta a una pregunta del cuestionario. Se le llaman comúnmente variables.

Records: es un grupo de variables que están lógicamente relacionadas.

Cuestionario: registros distintos con la misma identificación que forman un cuestionario.

Tipo de registro: si los registros en el archivo tienen la misma estructura, entonces se tendrán un solo tipo de registros en el archivo. Un archivo compuesto por un solo tipo de registro es llamado archivo plano. Si existen diferentes estructuras en los registros de un archivo, por ejemplo de población y vivienda, van a haber dos a más tipos de registro.

Un archivo de este tipo se le llama jerárquico. IMPS sólo puede procesar archivos secuenciales de texto, no jerárquicos.

Valor de NO aplicable: es el valor que la variable recibirá en el caso de que no posea ningún valor como consecuencia de alguna restricción. Por ejemplo, la fecundidad sólo es para mujeres. El valor común para indicar “no aplicable” es el espacio en blanco.

Valor de NO respuesta: es el código dado a una variable cuando una respuesta adecuada no es proporcionada. El valor común utilizado es el número 9 repetido en cada posición del campo.

Campos de identificación: estos campos identifican dónde se recabaron los datos de campo del cuestionario y también lo que identifica un cuestionario.

Largo de campo: es determinado de tal forma que la variable pueda tener espacio suficiente para contener los códigos o categorías de la variable. Por ejemplo, la variable “sexo” necesitará solamente una posición, pero la variable “edad” puede necesitar hasta 3 posiciones.

Contenido del campo: IMPS trabaja solamente con información numérica o alfanumérica. Se asume que la información está posicionada o justificada a la derecha en un campo declarado como numérico. En los campos declarados como alfanuméricos, la información se supone que está justificada a la izquierda.

Nombre de las variables: puede tener 1 a 16 letras, número y guiones. El primer carácter debe de ser una letra, el último debe de ser una letra o número. No es sensible al contexto (mayúsculas y minúsculas son iguales).

Subítem: una variable puede estar dividida en subvariables. Por ejemplo, la fecha puede consistir en tres variables: día, mes y año.

Punto decimal: IMPS trabaja sólo con datos enteros. El punto decimal está implícito y debe ser manejado por el programador. Para campos con decimales, el programa de captura mostrará el punto decimal al operador de captura, pero la información será almacenada como un número entero.

Valores aceptables, mínimo y máximo: durante la definición del diccionario de datos, se deben indicar los rangos de valores aceptables para cada variable. Al momento de la captura de datos, se realizará esta verificación y si la información ingresada no está en los límites aceptables, el operador recibirá un mensaje para su intervención.

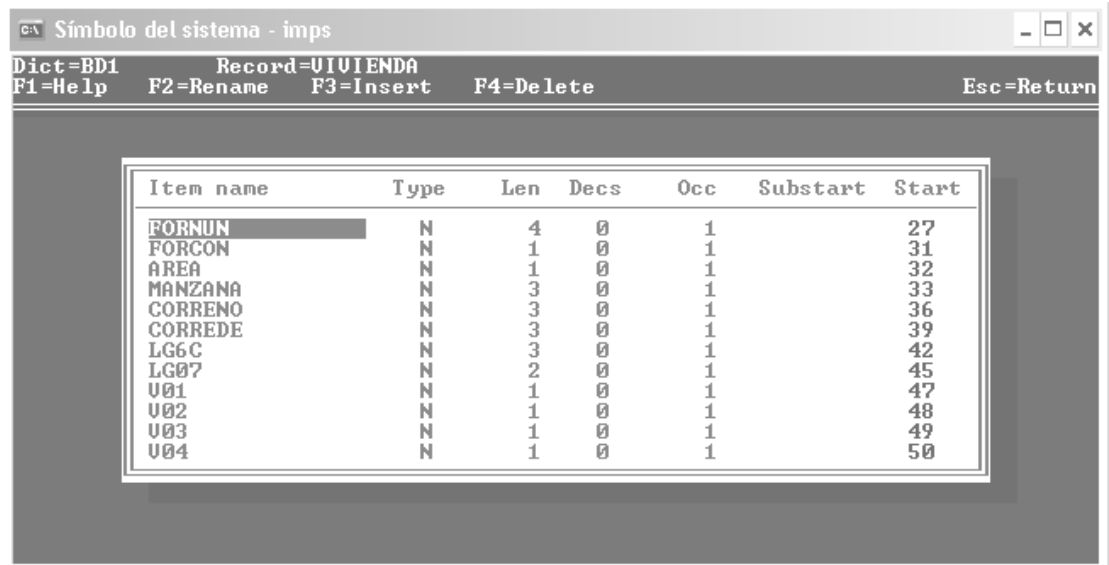
Categorías de la variable y nombres de categoría: cada variable puede tener varias categorías. Por ejemplo, “estado civil” puede tener (1) soltero (2) divorciado (3) casado y (4) viudo, espacio para no aplicable y (9) para no informo. Estas categorías son útiles ya que serán utilizadas para validar la captura y para los tabulados simples.

Redefinición de datos: puede ser útil redefinir datos de entrada de acuerdo con una distribución diferente. Por ejemplo, la variable “edad” será definida con un mínimo posible de 0 y un máximo de 99. Para producir tabulados simples en grupos quinquenales de edad, será necesario redefinir la variable como “edad05” dando una categoría individual para cada grupo de edades.

Variabes comunes: son aquellas variables que están en todos los registros del archivo.

Para definir variables de registros la pantalla es la siguiente:

Figura 2. Definición de diccionario de datos en IMPS



Item name	Type	Len	Decs	Occ	Substart	Start
FORNUM	N	4	0	1		27
FORCON	N	1	0	1		31
AREA	N	1	0	1		32
MANZANA	N	3	0	1		33
CORRENO	N	3	0	1		36
CORREDE	N	3	0	1		39
LG6C	N	3	0	1		42
LG07	N	2	0	1		45
U01	N	1	0	1		47
U02	N	1	0	1		48
U03	N	1	0	1		49
U04	N	1	0	1		50

Data Entry Centry (Entrada de datos)

Es un Módulo para diseñar pantallas de entrada orientadas por un diccionario. Al identificar la variable que se va a capturar, se pueden indicar saltos específicos a otras variables según el valor capturado.

CENTRY tiene características especiales de captura rápida; por ejemplo, al llenar completamente el campo digitado, automáticamente se salta al siguiente campo por digitar. También, si se desea, se puede solicitar que pare al digitador cuando ha ingresado un campo con valor incorrecto (basado en el diccionario). Al modificar una variable en el diccionario sólo es necesario recompilar la aplicación de entrada para que los cambios queden aplicados.

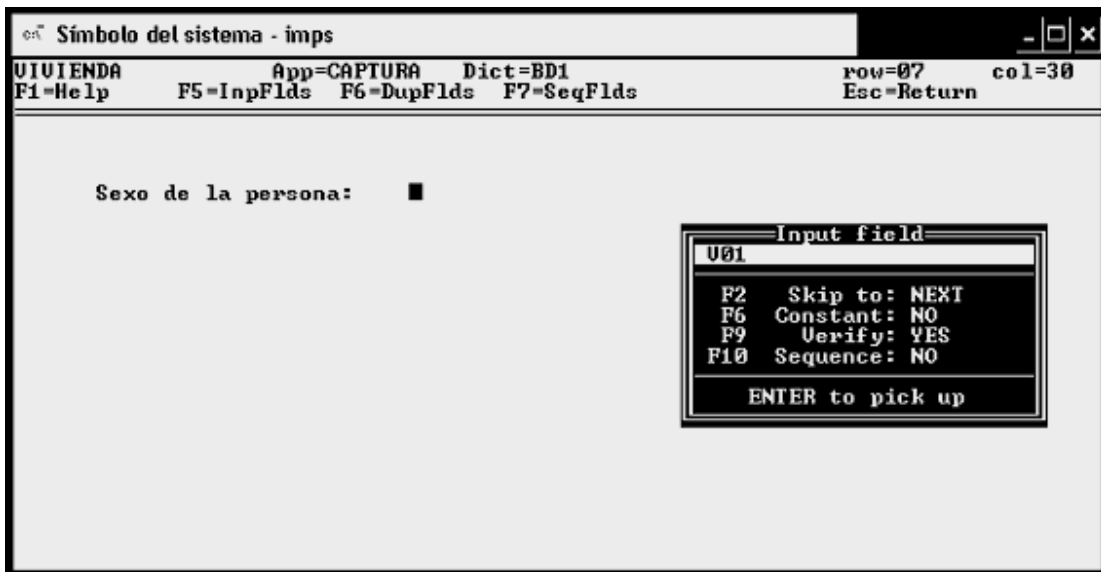
Cuenta con tres modos de operar:

- Ingreso de dato: se utiliza para la captura inicial de las boletas,
- Modificación de dato: sirve para ingresar a un cuestionario y modificar algún dato.

- Verificación de dato: es un re-ingreso del dato donde hace comparaciones con el primer ingreso y marca error donde hay diferencia. Cuando hay diferencia, solicita que se reingrese el dato y hace las correcciones necesarias dentro del archivo.

Una característica especial de CENTRY es que lleva controles estadísticos del digitador. Cada vez que el digitador ingresa, registra el modo de trabajo, tiempo transcurrido, teclados por minuto, ingresos incorrectos (según diccionario) y si es modo verificación registra errores tanto del digitador como el verificador. Estas estadísticas se guardan en archivos planos para poder ser administrados en sistemas de apoyo.

Figura 3. Ingreso de datos en IMPS



Consistencia del dato (CONCOR)

Es un Módulo sencillo y funcional para la detección de los errores del dato. Ya que utiliza los diccionarios como base, es muy sencillo utilizar las técnicas existentes (estáticas y dinámicas) de consistencia e imputación del dato.

Algunas de las funciones de CONCOR son:

- Verificar la estructura del cuestionario.
- Verificar los rangos de las variables.
- Verificar la consistencia dentro del registro y entre los registros.
- Corrección automática.
- Generar archivos de salida diferentes al de entrada, incluso en su estructura
- Generar informes de cómo los errores fueron corregidos, lo cual es muy útil para revisar el sesgo que ha sufrido la información
- Ser utilizado conjuntamente con CENTRY para que las correcciones interactivas puedan ser rechazadas.
- Permite hacer “file matching” o pareo de datos con otro archivo.
- Tiene la facilidad de definir matrices de una manera sencilla y gráfica, que logra que sea sencillo utilizar técnicas dinámicas de consistencia.
- Utiliza como entrada sólo archivos secuenciales en formato ASCII, lee un cuestionario y lo guarda en memoria y permite al programador hacer referencia a sus registros y variables para programar las diferentes pautas para dar consistencia al dato. Es conveniente que el archivo se encuentre clasificado por área de recolección del dato debido a que las imputaciones pueden variar de un lugar a otro.

El lenguaje que utiliza cuenta con definición de variables numéricas (medium, large, xlarge, alpha) y con las instrucciones siguientes:

- Condicionales: if, universe
- Asignaciones: impute, let, increase
- Bucles: do, find, count, sum
- Recodificaciones: recode, crecode
- Salida: output, write
- Control de flujo: stop, call
- Generador de informes: messages, write

- Para rangos, se pueden indicar de manera sencilla, ejemplo: 5:12 toma los datos 5,6,7,8,9,10,11 y 12 o para comparar alguna variable con valores se puede utilizar de la manera siguiente: 5,9,12.

Ejemplo

```
If P17-Relación<>1,3 6:8  
  
    Let V01=P17B  
  
Else  
  
    MSG "ERROR 66: Relación inválida"  
  
    Call SUBE66  
  
End-if
```

Se permite incluir instrucciones en lenguaje COBOL que no van a ser interpretadas por CONCOR, sino que serán incluidas directamente en el programa generado por CONCOR. De haber errores sintácticos, sólo serán detectados durante la compilación efectuada en REALIA COBOL.

Se pueden incluir instrucciones de definición de datos o instrucciones ejecutables que serán incluidas directamente en *Working-Storage Section* y *Procedure Division* del programa de cobol.

Tabulación de datos (CENTS)

Es un Módulo que se utiliza para generar tabulados simples y complejos. La idea básica de CENTS es dibujar los cuadros, definiendo las filas y columnas, y luego elaborar un programa que cuente, según la condición, en su respectiva fila y columna. Algunas de sus características son:

- SubMódulo para diseño de los cuadros donde se definen las filas y columnas que se utilizarán y se realizan los cálculos de totales, promedios, desviaciones, etc. sobre las columnas y/o filas. Además, se pueden definir niveles jerárquicos para producir los cuadros.
- Lenguaje de programación similar a CONCOR con instrucciones para contar por fila y columna (TALLY (row, col)).
- Programa para consolidar datos a diferentes niveles jerárquicos de presentación.

Por ejemplo, para realizar el siguiente tabular:

	Total	Hombres	Mujeres
Guatemala			
Guatemala			
San Miguel Petapa			
El Progreso			
Guastatoya			
Morazán			

Únicamente se requiere dibujar el formato de la siguiente manera:

Depto. /Mupio	Total	Hombre	Mujeres
Are Text	xxxxxxxx1	xxxxxxxx2	xxxxxxxx3

realizar el calculo ROWSUM 1, 2-3

y escribir el siguiente programa:

Área-estructura Departamento Municipio

Table C01(1,3)

For each(Persona)

Recode Sexo to Row

1,2 Others,3

Tally C01(1,Row) y End-for

Basta eso para producir un cuadro con todos los departamentos y municipios.

Figura 4. Pantalla de dibujo de tabulados

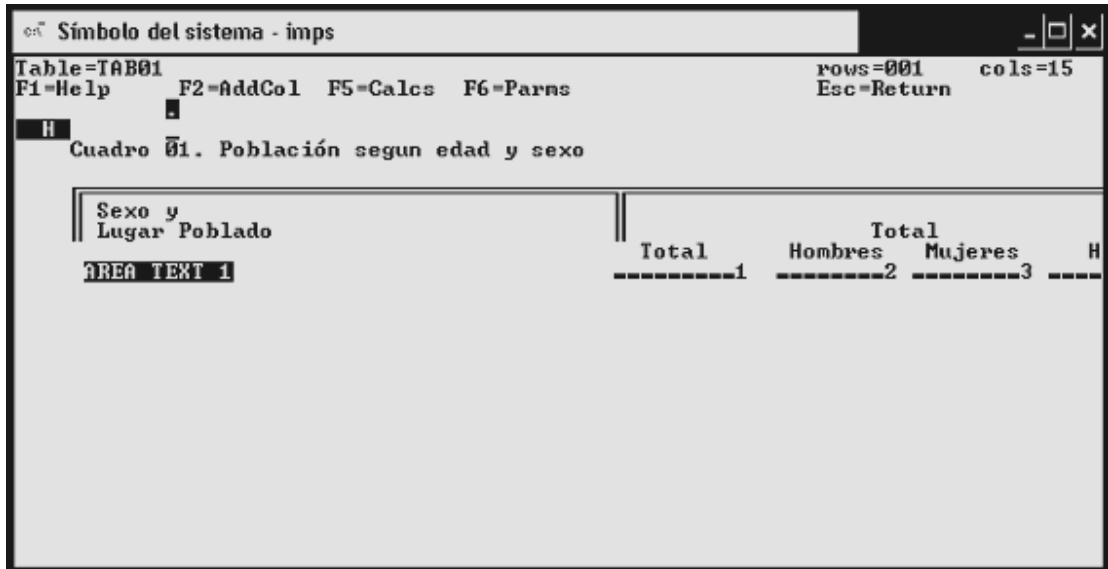
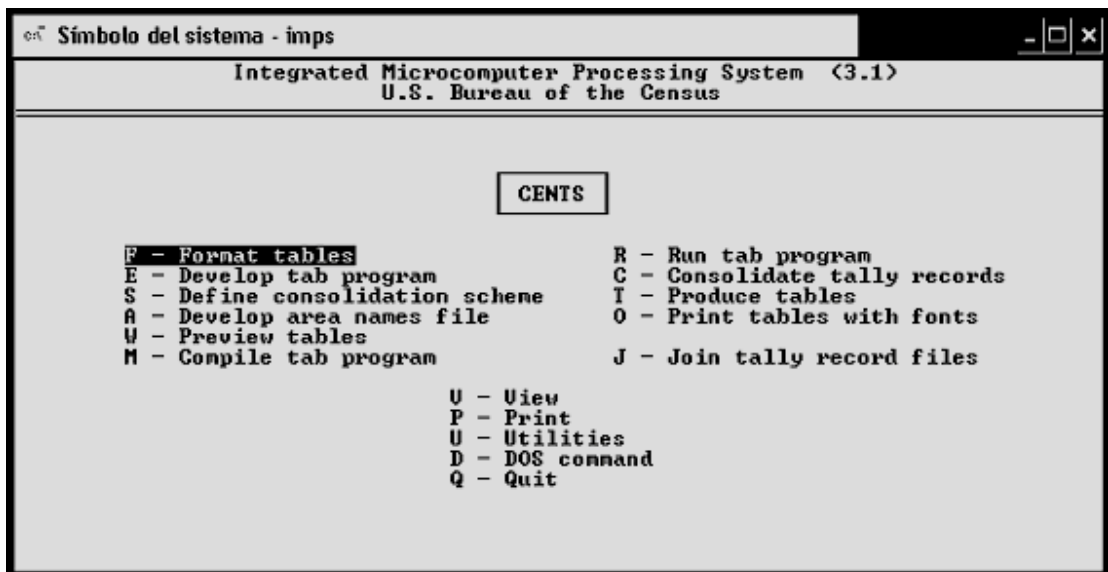


Figura 5. Menu de CENTS



3.2.2 CPro

El sistema Census and Survey Processing System (CPro) (Sistema de procesamiento de datos para censos y encuestas) es un *software* que cuenta con las características necesarias para procesar un censo: definición de diccionario, entrada de datos, consistencia y tabulaciones. CPro combina las facilidades de IMPS y las ventajas de ISSA (otro lenguaje de procesamiento de datos para sistemas DOS) en un ambiente gráfico amigable.

CPro permite crear, modificar y correr una aplicación, darle consistencia, tabularla de una manera sencilla y en un ambiente de trabajo integrado. Procesa archivos con las características de IMPS con uno o más tipos de registros. Al ingresar archivos a través de CPro, los guarda en archivos de texto descritos por diccionarios. CPro contiene un conjunto de procedimiento y lenguaje para implementar el control y reglas de ingreso de datos.

CPro también provee herramientas para visualizar archivos de texto, ver mapas temáticos creados por CPro, conversiones de datos para IMPS y lectura de mapas elaborados en ERSI.

CPro está siendo elaborado y mejorado por el Buró de Censos de Estados Unidos y una empresa de desarrollo, y se puede utilizar y distribuir gratuitamente.

3.2.2.1 Componentes generales de CPro

3.2.2.1.1 Diccionario de datos

En este Módulo se describe la organización del archivo de datos. Esta descripción se utiliza en todos los demás módulos de CPro en orden de acceso y uso en las diferentes aplicaciones.

El diccionario provee etiquetas para la entrada de datos, así como conjuntos de valores para utilizar en los tabulados y rangos permitidos de entrada. A diferencia de IMPS, aquí se incorporan los conceptos de niveles jerárquicos que pueden contener uno o varios tipos de registros. Normalmente para un censo de población o agropecuario sólo es necesario un nivel, pero para cuando se necesite representar una estructura jerárquica es permitido hasta tres niveles.

Además, el diccionario se utiliza para suministrar una descripción o foto de cómo están o estarán los datos guardados en la computadora. En CPro puede estar el punto decimal como parte del archivo o implícito como en IMPS.

El diccionario de datos facilita la comunicación, provee una manera sencilla de utilizar los datos y que cualquier programador pueda utilizar un archivo de un censo o encuesta.

Figura 6. Diccionario de datos en CPro

Item Label	Item Name	Start	Len	Data Type	Item Type	Occ	Dec	Dec Char	Zero Fill
(record type)		1	1	Alpha					
DEPTO	DEPTO	2	2	Num	Item	1	0	No	Yes
MUPIO	MUPIO	4	2	Num	Item	1	0	No	Yes
SECCION	SECCION	6	3	Num	Item	1	0	No	Yes
SECTOR	SECTOR	9	1	Num	Item	1	0	No	Yes
NUMHOG	NUMHOG	10	3	Num	Item	1	0	No	Yes
DEPTOGE0	DEPTOGE0	14	2	Num	Item	1	0	No	Yes
MUPIOGE0	MUPIOGE0	16	2	Num	Item	1	0	No	Yes
FORNUM	FORNUM	18	4	Num	Item	1	0	No	Yes
FORCON	FORCON	22	1	Num	Item	1	0	No	Yes
AREA	AREA	23	1	Num	Item	1	0	No	Yes
MANZANA	MANZANA	24	3	Num	Item	1	0	No	Yes
CORRENO	CORRENO	27	3	Num	Item	1	0	No	Yes
CORREDE	CORREDE	30	3	Num	Item	1	0	No	Yes
LG6C	LG6C	33	3	Num	Item	1	0	No	Yes
LG07	LG07	36	2	Num	Item	1	0	No	Yes
V01	V01	38	1	Num	Item	1	0	No	Yes
V02	V02	39	1	Num	Item	1	0	No	Yes
V03	V03	40	1	Num	Item	1	0	No	Yes
V04	V04	41	1	Num	Item	1	0	No	Yes
V05	V05	42	1	Num	Item	1	0	No	Yes
IH01	IH01	43	1	Num	Item	1	0	No	Yes
IH02	IH02	44	1	Num	Item	1	0	No	Yes
IH03	IH03	45	1	Num	Item	1	0	No	Yes

3.2.2.1.2 Entrada de datos

Módulo para agregar, modificar y verificar cuestionarios. Con este Módulo se pueden crear una ilimitada cantidad de ventanas o formas para captura en ambiente gráfico basándose en diccionarios definidos. Crea automáticamente movimientos de pantalla hacia la derecha, izquierda, arriba o abajo según sea necesario.

Se pueden elaborar chequeos de consistencia, saltos condicionales y programación por campo cuando se llega y cuando se sale de él, lo cual permite una flexibilidad excelente de programación para la captura de datos.

Se tiene que tomar en cuenta que la programación depende del método que se utiliza para ingreso: a) Tecleo de cabeza abajo: éste se utiliza principalmente en volúmenes grandes de información como censos; en este método se debe utilizar la menor programación para que el ingreso sea rápido y permitir a los programas de consistencia la revisión de los datos; b.) Tecleo de cabeza arriba: éste se puede utilizar en encuestas donde los cuestionarios son grandes y complejos por lo que requiere mayor control en el ingreso.

CSEntry ofrece los dos sistemas de captura: controlada por operador o por sistema, la diferencia de cada una de ellas es la manipulación que puede hacer el digitador sobre la aplicación. Al hacerlo por sistema se requiere de un archivo de parámetros donde se indica el *path* para guardar los datos y otros parámetros necesarios los cuales permanecerán inamovibles en la captura.

Se pueden mostrar mensajes con respuesta, utilizar combos de opciones para ingresos, cuadrículas de trabajo, acceso a múltiples archivos para realizar revisiones de datos (en tablas auxiliares), lectura sobre archivos secundarios y otras ventajas que lo hacen más poderoso y dinámico que CENTRY de IMPS.

Tiene integrado un sistema de estadísticas para control de los digitadores, el cual permite llevar rendimientos de producción y calidad del dato.

Figura 7. Ingreso de datos en CPro

3.2.2.1.3 Consistencia del dato

En CPro se pueden desarrollar programas BATCH que corren sobre archivos de texto y diccionario relacionado. Aquí se puede utilizar un poderoso lenguaje de CPro para programar consistencias estáticas y dinámicas.

Cada variable dentro del diccionario puede tener su propio procedimiento para revisar la consistencia y también se pueden generar procedimientos independientes.

A continuación se muestra un ejemplo de imputación estática:

```
PROC GLOBAL
```

```
toggle_sex = 1;
```

```
PROC SEX
```

```
if $ = notappl or not ($ in 1:2) then
```

```
    $ = toggle_sex;
```

```
    toggle_sex = 3 - toggle_sex;
```

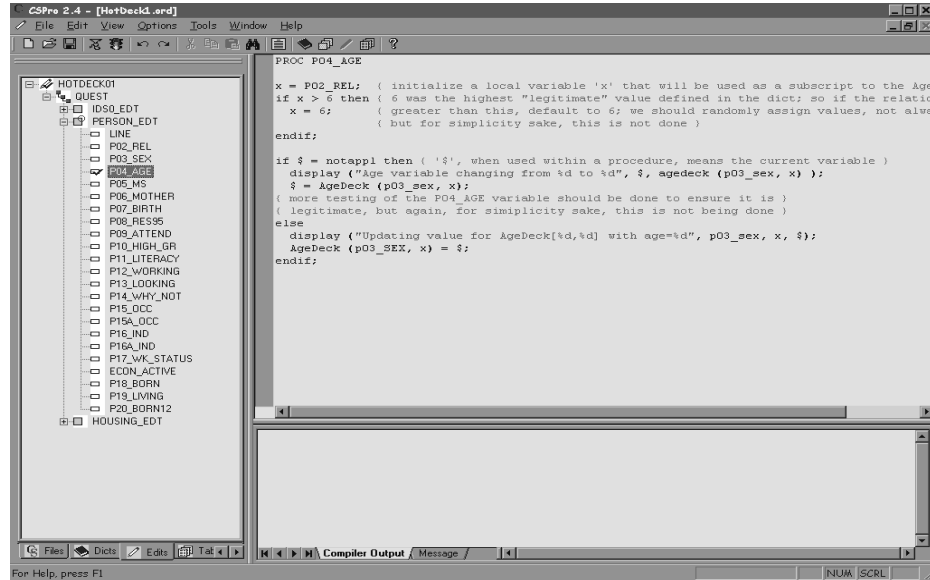
```
endif
```

Además, tiene la facilidad de construir matrices fácilmente. Ejemplo de definición de matriz dinámica.

	(Jefe)	(Esposa)	(Hijo)	(Abuelo)	(Otro)	
	1	2	3	4	5	
(M)	1	35	50	10	41	65
(F)	2	32	48	10	37	68

CSPPro tiene la facilidad de comunicación con los archivos a través de los diccionarios y el poder de un lenguaje, que, combinados, pueden generar excelentes programas para detección y corrección de errores de los registros y entre los registros.

Figura 8. Editor de programas para consistencia



3.2.2.1.4 Tabulación

CSPRO cuenta con un Módulo para realizar cruces rápidos de variables y frecuencias. Los cruces rápidos que genera tienen un máximo de 2x2 variables y se pueden incluir cortes jerárquicos, como departamentos, municipios, área urbana y rural, etc.

La generación de estos tabulados es sencilla, sólo se arrastran las variables del diccionario hacia la hoja de tabulados y listo.

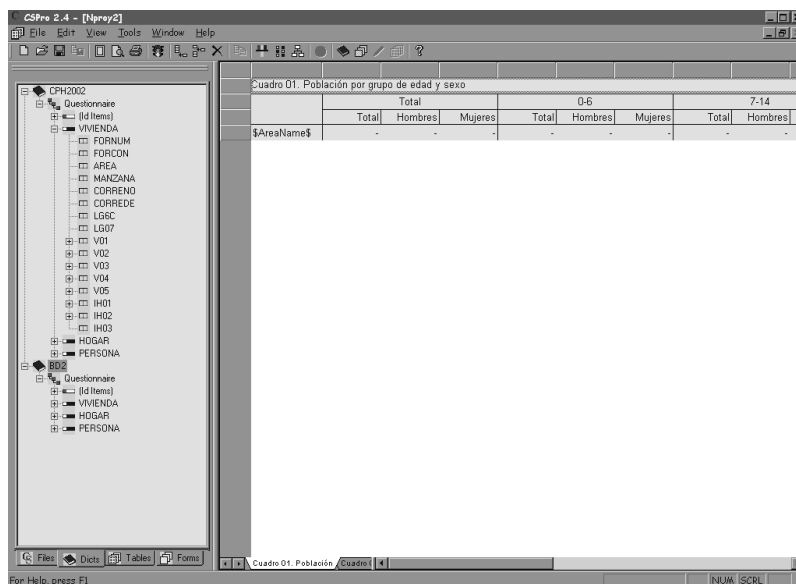
Se pueden incluir universos, condiciones, valores indefinidos, valor para expandir el dato en el caso de una encuesta y otras ventajas que permiten la facilidad del análisis del dato.

Aunque es muy poderoso, todavía se está desarrollando la versión para generar cruces complejos.

Incluye una opción para generar mapas temáticos basados en los cruces elaborados. Para desarrollar estos mapas temáticos se debe contar con los dibujos de los mapas en los diferentes niveles de división geopolítica en herramientas como ESRI.

En general, CPro es un *software* sencillo de utilizar y con una ayuda completa que los programadores pueden utilizar para familiarizarse con sus bondades.

Figura 9. Módulo para cruces rápidos de variables



3.3 Software de reconocimiento de datos

3.3.1 Evolución de procesamiento de datos

El procesamiento de los censos ha experimentado cambios profundos a partir del ingreso de las computadoras en el campo del reconocimiento óptico. Antes de 1947 los censos de población se procesaban manualmente. Procesar manualmente un censo significaba contar con muchas personas para contabilizar las respuestas una a una –con los clásicos “palotes”-, e ir haciendo sumatorias parciales hasta llegar a las cifras básicas.

En esos años se introdujo una primitiva máquina lectora que detectaba la existencia de perforaciones practicadas sobre la superficie de una tarjeta. El significado que se le asignaba a las perforaciones contabilizadas en la misma posición se establecía posteriormente en una etapa de interpretación que ponía en correspondencia cada pregunta con el conjunto de sus respuestas posibles. Podemos decir que los cuestionarios recibían un tratamiento similar al de las antiguas boletas de quiniela.

Paulatinamente la informática asume los sucesivos descubrimientos en la materia, desarrollando herramientas para reconocer marcas, caracteres y códigos de barra. Ya a partir de 1980 se han utilizado lectoras de marcas, es decir, un sistema que puede detectar casilleros rellenos y vacíos. Esa lectora solo podía procesar preguntas con respuestas predeterminadas (cerradas). La persona que llena la boleta con respuestas posibles lo hace relleno con lápiz de grafito la “burbuja” correspondiente al dato que proporcionaba el entrevistado.

Por primera vez el reconocimiento estaba asociado a un programa (OMR: Optical Mark Reading *software*, programa de detección óptica de marcas) que interpretaba qué significado tenía una marca en determinado lugar del cuestionario. Por lo tanto, la interpretación de respuestas dejó de ser la segunda etapa del procesamiento y se integró a las funciones de la lectora.

De todos modos, a pesar de este avance, las preguntas de escritura abierta se trataban en una etapa posterior, requiriendo de muchas personas que analizaran e interpretaran lo que se había consignado en campo y luego se tenía que poseer un equipo de digitadores para capturar los códigos de las preguntas abiertas.

Recién entonces comenzaba el procesamiento. Debido a las deserciones que se produjeron y a los problemas que implicaban las nuevas incorporaciones de personal, los tiempos se extendían interminablemente y así fue como, en los años 90, el procesamiento de esos campos temáticos todavía llevaba tiempo de casi 3 a 4 años.

Para los años 2000 se procesan censos con desarrollos últimos de captura óptica de datos. Además del sistema de reconocimiento de marcas que se ha empleado en censos anteriores, se utilizan nuevos métodos: OCR (*Optical Character Recognition software*, programa de reconocimiento de caracteres) para reconocer caracteres impresos, e ICR (*Intelligent Character Recognition software*) para manuscritos. Este último es particularmente útil para codificar y procesar las preguntas abiertas.

3.3.2 Descripción de tecnología

Para la ejecución del procesamiento de datos se ha estado utilizando la tecnología de reconocimiento de marcas (OMR), reconocimiento de caracteres (OCR) y reconocimiento inteligente de caracteres (ICR). A continuación se hace una comparación de tecnologías para tener una idea clara de ellas.

3.3.2.1 ICR

La tecnología ICR provee sistemas de escaneo de imágenes con la habilidad de convertir caracteres escritos a mano (no tipo carta) en caracteres legibles en un computador.

Las imágenes de los caracteres escritos a mano se extractan de un BMP (bitmap) de la imagen escaneada. Esto incluye el escáner de formularios para la interpretación de lo que se ha ingresado.

3.3.2.2 OCR

La tecnología OCR provee sistemas de escaneo de imágenes con la habilidad de convertir caracteres escritos a máquina o en computador a caracteres interpretables en un programa de computación. Las imágenes de los caracteres escritos a máquina se extractan de un BMP (bitmap) de la imagen escaneada. Esto incluye el escáner de formularios para la interpretación de lo que se ha ingresado.

3.3.2.3 OMR

La tecnología OMR detecta la ausencia o presencia de una marca o seña en un lugar determinado sin incluir la figura o texto escrito. La mayoría de los *software* interpretan la posición de la marca en el escáner y lo traducen a código ASCII deseado. También incluye el escaneo de formularios y su interpretación. En los formularios existen pequeños círculos (marcas de selección) a los cuales se les llama “burbujas” que llena el empadronador según sea el caso y luego se interpretan como campos. Esta tecnología no reconoce los caracteres escritos a mano o a máquina. OMR no crea una imagen de la forma.

Una producción media de scaneo de imágenes va entre 1,200 y 7,500 formularios por hora.

Pero de cualquier manera no es la velocidad de escaneo la que determina la producción de datos sino otros factores, como los números de caracteres por página, diferentes tipos de marcas, tipos y legibilidad de letra a mano. Estos son los que afectan la producción de datos.

La tecnología ICR y OCR es significativamente más cara que un OMR. Dependiendo de la cantidad de formularios escaneados, mayor será la cantidad necesaria de servidores y estaciones de trabajo para procesar esta actividad. El embudo de la información raras veces será por el escaneo de imágenes: por el contrario será determinada por la capacidad instalada de la red.

3.3.2.4 Comparaciones

Las diferencias entre ICR, OCR y OMR son demostradas en las siguientes áreas:

- Formas

- Almacenamiento y recuperación
- Exactitud.

3.3.2.4.1 Formas

Los formularios OMR son documentos especializados, y tienen marcas de control y tiempo para saber dónde están cada una de las marcas que se deben reconocer. También contienen marcas de identificación que lucen como cajas negras en los bordes superiores e inferiores del formulario. Adicionalmente, el corte de estos formularios es preciso y las burbujas deben estar localizadas exactamente en el mismo lugar en cada formulario.

La tecnología ICR y OCR es más complaciente ya que no tiene marcas de tiempo ni formas de identificación por lo que las imágenes pueden hasta cierto grado moverse. Estas tecnologías usan marcas de registro que a menudo aparecen como triángulos en las cuatro esquinas de un formulario para reconocimiento de las ubicaciones de las imágenes.

Los documentos que tendrán información extractada y enviadas al *software* de ICR contienen cajas impresas en el formulario que los empadronadores deben llenar (letra por letra).

En ambos tipos de documentos el color es muy importante. El uso de color de imprentas especializadas reduce el problema de transmisión de imágenes a las computadoras. En OMR se permiten lapiceros y lápices para completar el formulario; sin embargo, para ICR y OCR se deben utilizar lápices o lapiceros que cumplan con lo requerido. En los diferentes tipos de tecnología, el color y tipo de papel determina la exactitud de recuperación.

3.3.2.4.2 Almacenamiento y recuperación

Si los documentos serán almacenados electrónicamente (base de datos), la imagen es necesaria. Con ICR y OCR la imagen puede ser escaneada, ordenada y escrita en cualquier medio óptico. Las imágenes pueden ser recuperadas e impresas como se necesiten. Con OMR las imágenes o documentos no se leen por escáner por lo que no es posible su almacenamiento y recuperación.

3.3.2.4.3 Exactitud

Con un documento impreso con calidad, OMR puede dar el 99% en la exactitud de lectura de datos. ICR y OCR puede proveer el 98% de exactitud si el sistema está apropiadamente parametrizado; los formularios deben estar bien diseñados, los caracteres del empadronador deben estar claros y nítidos, y se usa la edición contextual.

Esta es aproximadamente la misma exactitud obtenida por la captura tradicional de un digitador de datos.

3.3.2.5 Tabla de comparaciones

	ICR-OCR	OMR
Reconocimiento de escritura a mano	Sí –ICR	No
Reconocimiento de escritura a máquina	Sí –OCR	No
Reconocimiento de marcas tipo burbuja	Sí	Sí
Reconocimiento de marcas tipo “X”	Sí	No
Reconocimiento de códigos de barra	Sí	Sí
Requiere marcas de tiempo en formulario e ID’s de formas	No	Sí
Requiere marcas de registro	Sí	No
Exactitud	Arriba de 98% con altibajos	Consistente en 99.9%
Almacenamiento y recuperación electrónica	Sí	No
Ingreso por FAX	Sí	No
Velocidad	Sin límite – Depende de la configuración de la red y cantidad de escáner	1,500 – 10,000 HR/escáner

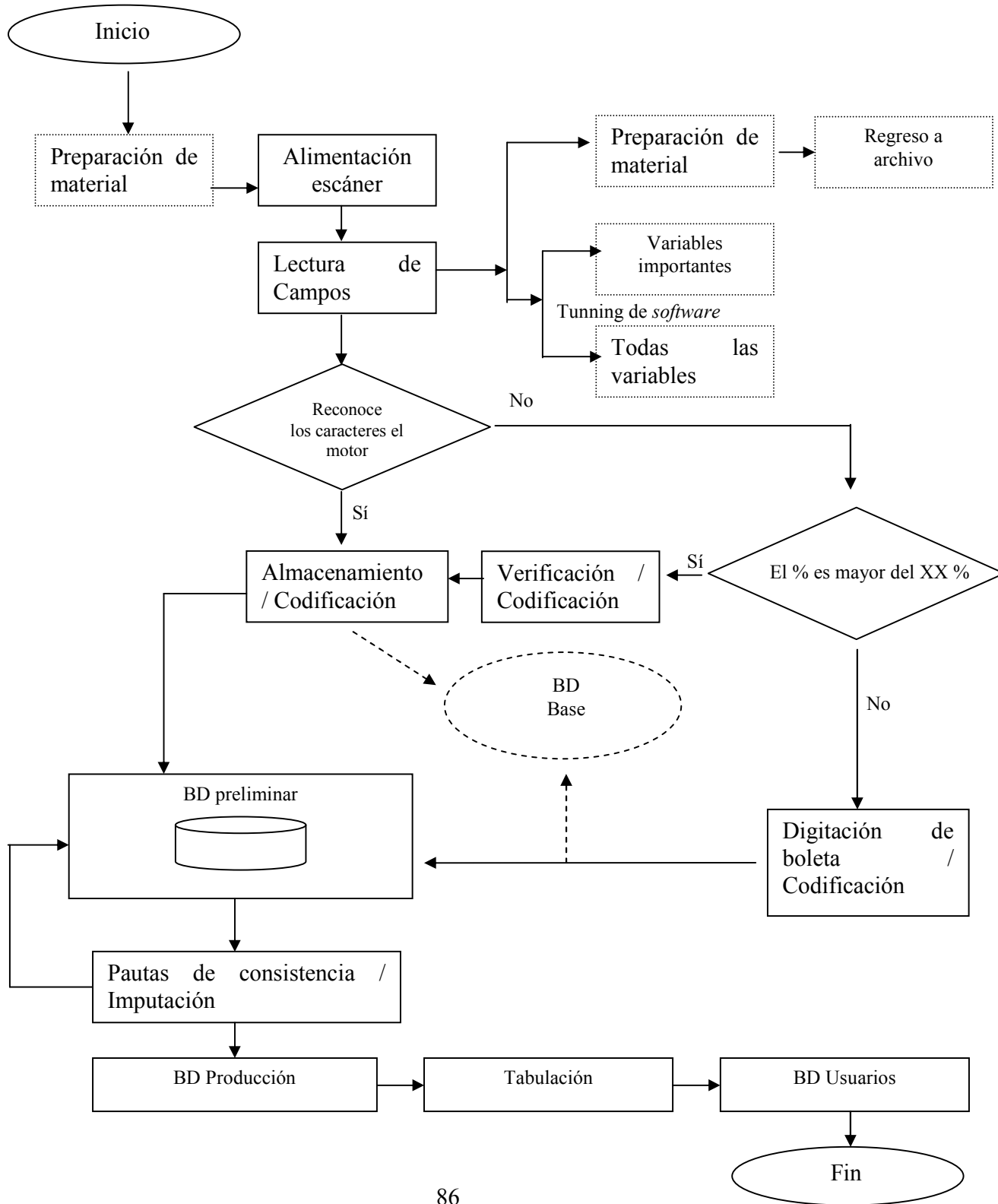
3.3.2.6 Requerimientos

Utilizando la tecnología anteriormente descrita, se ha desarrollado *software* que elimina la digitación por teclado. Para poder implementar este tipo de tecnología en procesamiento de datos censales y de encuestas se requiere de por lo menos los siguientes componentes:

- Escáner de alta velocidad
- Administrador de imágenes
- Motores de reconocimiento de caracteres
- Sistema de revisión de datos no reconocidos
- Mecanismos de control de dato
- Herramientas de exportación de archivo

3.3.3 Diagrama de flujo para procesamiento óptico

Figura 10. Diagrama de flujo para procesamiento óptico



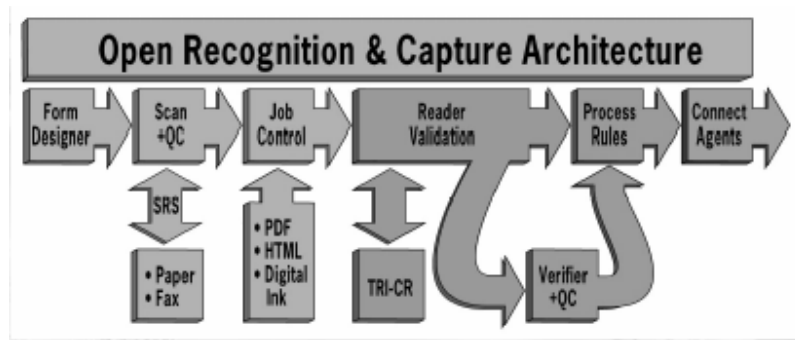
3.3.4 Teleform

Teleform es un *software* desarrollado por la empresa *Cardiff*, que posee la tecnología de reconocimiento de imágenes. Teleform está disponible para negocios y organizaciones de gobierno que quieren tener un sistema rápido y fácil de captura y conversión de formas en papel, creadas por el mismo *software*, a datos digitales que pueden ser utilizados y explotados.

Teleform ofrece una solución para reducir los costos de operación asociados a la digitación de todos los datos manualmente, estimando que disminuye el tiempo de respuesta, incrementa la calidad del dato y acelera la presentación de los resultados.

Este *software* utiliza el concepto de *batch*, que es un conjunto de cuestionarios de campo con la misma identificación, los cuales escanea y almacena. Una vez almacenada la información, pasa por el proceso de reconocimiento que utiliza 5 motores los cuales son como votantes para determinar que caracter es el que se está reconociendo. Si tres de los votantes reconocen el caracter, es aceptado; de lo contrario, se almacena en repositorio para ser verificado por personas que hacen de digitadores. A ellos se les activa un menú con los *batch* pendientes de revisar y que muestra las imágenes (cuestionarios escaneados) y pide la confirmación del carácter con duda.

Figura 11. Diagrama de procesos de Teleform



Dentro de Teleform se encuentran los siguientes módulos:

Creador de formas: con este Módulo se diseñan las boletas que serán utilizadas para la recolección de datos en campo. Además, digitaliza y configura formas existentes, define características y especificaciones de las formas, define opciones de exportación y verificación de los datos. Aquí se pueden escribir *scripts* de programas en Visual Basic y otras herramientas de desarrollo. Es importante mencionar que al definir la forma, se pueden hacer enlaces con tablas descriptivas de códigos, establecer rangos permitidos y hacer codificaciones de preguntas abiertas de manera asistida.

Figura 12. Creador de formas de teleform

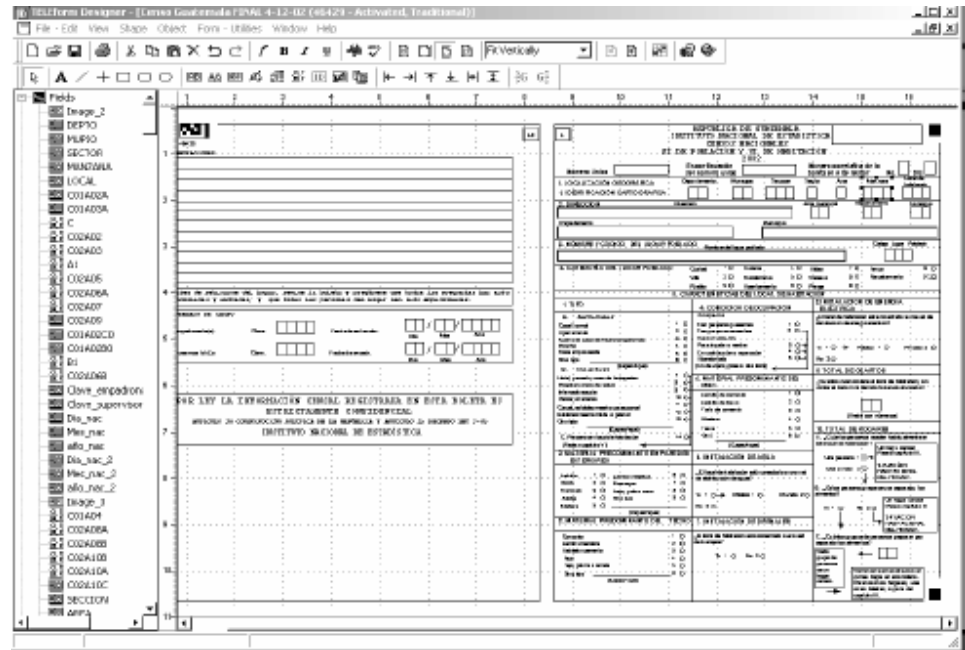


Figura 13. Estación de escaneo y administrador de los batch

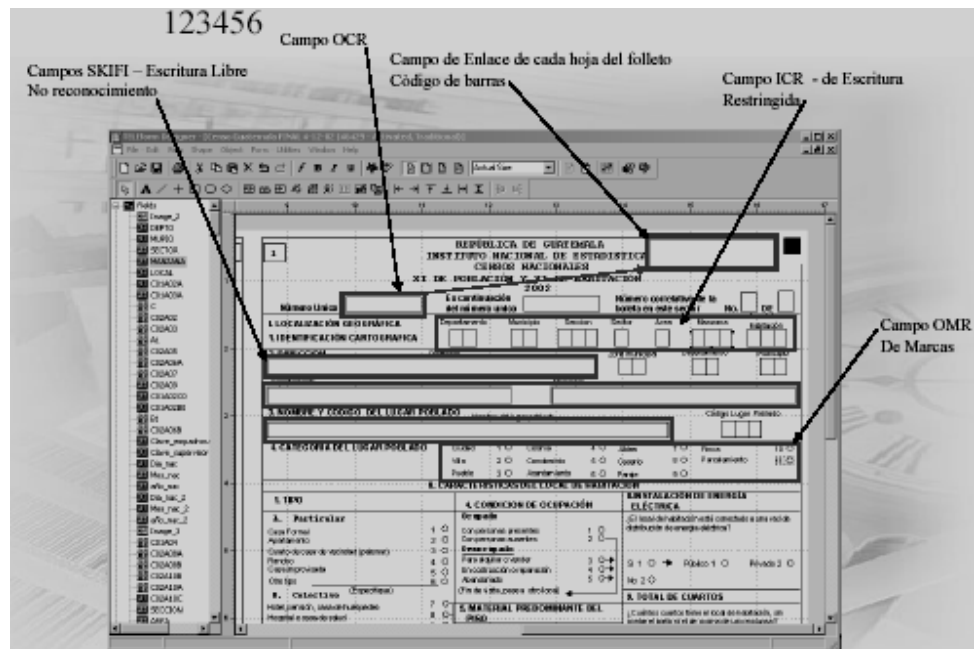
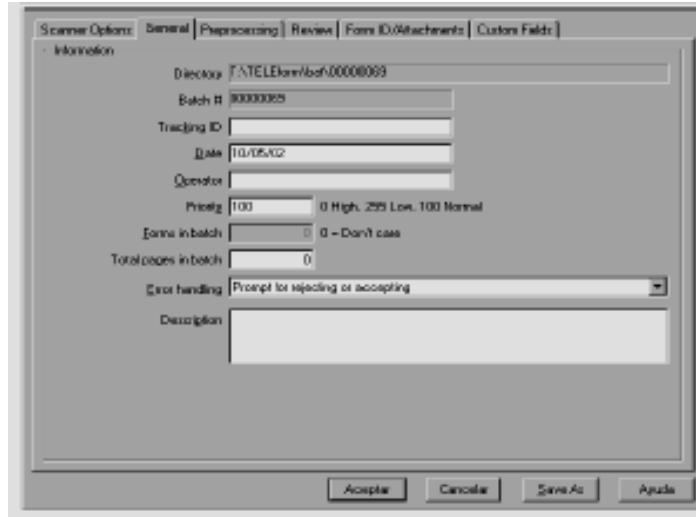


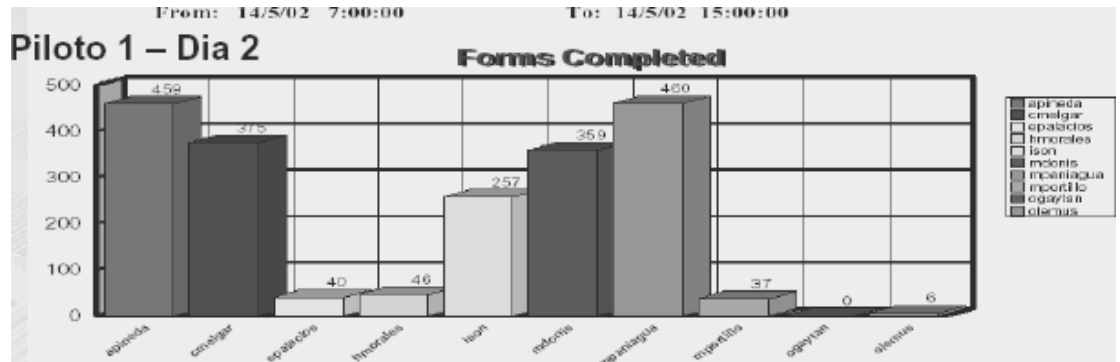
Figura 14: Administrador de proyecto



Control de proceso: lleva estadísticas sobre estatus de los *batch*, rendimientos de los verificados, grado de reconocimiento y otras estadísticas vitales para controlar el proceso.

Se pueden extraer reportes como: a) rendimiento de sistema: este reporte ayuda en el análisis de la eficiencia operativa del sistema Teleform, b) resumen de evaluaciones de las formas: provee información del tiempo que toman las formas para ser evaluadas por el lector, c) corrección de formas: facilita identificar en qué modalidades de corrección los operadores sobresalen en productividad y en cuáles necesitan más practica y entrenamiento, d) detalle de las formas: este reporte puntualiza las ineficiencias en el diseño de la forma, e) rendimiento de operadores: le permite identificar la eficiencia de los operadores en la verificación, así como si están corrigiendo con eficiencia o necesitan atención y d) resumen de lotes: estadísticas útiles si se están procesando uno o más lotes y se desea evaluar el estado del proceso.

Figura 15. Estadísticas de procesos



Lector: sólo los motores que utilizan los algoritmos de reconocimiento. Es el responsable de interpretar el contenido de documentos digitalizados, faxes, imágenes y datos en formularios. Teleform utiliza el reconocimiento neurológico Tri-CR.

Figura 16. Definición de tipos de datos de Teleform

Tipos de Campos en el formulario

- Campo OCR:** Campo de reconocimiento de texto.
- Campo de Imagen Reconocimiento OCR:** Campo de reconocimiento de imágenes.
- Código de Barras:** Campo de reconocimiento de códigos de barras.
- Campos de escritura libre - No reconocimiento (SKIFI):** Campos de texto libre que no se reconocen automáticamente.
- Campos de escritura restringida - Reconocimiento Inteligente ICR:** Campos de texto con estructura predefinida que se reconocen automáticamente.
- Campos de Marcas - Reconocimiento Inteligente OMR:** Campos de selección de opciones (checkboxes, radio buttons) que se reconocen automáticamente.

A diferencia del OMR puro, estas no necesitan estar totalmente llenos, ya que al diseñar la forma, definimos el % necesario para considerar una marca como válida. También no necesitan de Timing Marks en el formulario.

Verificación de lectura: es el Módulo con el cual el verificador puede confirmar los datos que no se han podido reconocer. Se responsabiliza de mostrar la imagen de los documentos con la variable marcada por el lector de Teleform a fin de que el operador pueda corregir o confirmar el dato. Utiliza técnicas que imitan en forma rítmica y sincrónica, la forma en que la mente humana ve y procesa datos.

Figura 17. Verificación de lectura en el módulo con el cual se pueden confirmar los datos que no se han podido reconocer

Conectividad externa: es el componente con el cual Teleform genera los *batch* revisados y reconocidos a otras herramientas, por ejemplo: SQL, Oracle, Access, XML, ASCII, CSV y SPSS

Además de este *software*, existen otros en el mercado que varían en la cantidad y calidad de los motores, el autoaprendizaje y el modo de verificar los datos; entre ellos están TIS (TOP Image Systems) el cual es un *software* israelí con ventajas como autoaprendizaje de diferentes tipos de letra: contiene los mismos módulos básicos del TELEform. También está FormsPro (The Forms Processing System) que se parece mucho al TELEform e incluye el nuevo concepto de IFR (Intelligent Form Recognition) que reconoce automáticamente las formas definidas por el diseñador.

La selección del *software* depende mucho del apoyo y acompañamiento de la empresa que lo ofrezca ya que algunos de ellos necesitan *tuning* que lo pueden hacer únicamente los programadores de la casa matriz.

3.4 Comparación de *software* por teclado y reconocimiento óptico

Actualmente se utiliza la tecnología de reconocimiento óptico con fines de:

- Minimizar tiempos de captura tradicional de datos
- Mejorar la calidad de datos procesados
- Minimizar costos de procesamiento

Se ha comprobado que la tecnología óptica es funcional, sin embargo, para poder tomar la decisión es necesario tener en cuenta los siguientes aspectos:

El entorno de trabajo tiene mucho que ver: para que las boletas sean bien escaneadas es necesario que se mantengan limpias y en un ambiente físico fresco. Esto es complicado debido a la manipulación que tienen en campo dichos cuestionarios.

La recolección del dato: para tener éxito en el reconocimiento de caracteres se requiere utilizar cierto tipo de letra, marcas bien elaboradas en respuestas cerradas, mantenerse dentro de los límites de los cuadros de respuesta y otras especificaciones que son difíciles de mantener por el nivel académico de los empadronadores de campo. Se requeriría especializar a los empadronadores y eso llevaría tiempo y un costo alto.

Apoyo local de empresas: actualmente todavía no se cuenta con empresas que tengan la experiencia y soporte técnico local para apoyar los procesos masivos de datos.

Existe siempre el error de reconocimiento: aun con las mejoras de los motores de reconocimiento, todavía se confunden con diversidad de empadronadores. Cuando es pequeña la cantidad de cuestionarios y estos han sido elaborados por las mismas personas, los sistemas aprenden rápido y son eficientes; sin embargo, con muchos empadronadores, el programa le cuesta aprender y lo que aprende es difícil que le sirva para otro empadronador.

Poca contratación de personal: a pesar de que se requiere de intervención humana para revisar los caracteres no reconocidos, no se puede comparar con la cantidad de personal que se requiere para digitar tradicionalmente. Los costos que incluyen los escáners, *software*, equipo, soporte técnico y mantenimiento se equiparan con el de contratar muchos digitadores en diferentes turnos.

En nuestro país es necesario generar fuentes de trabajo por lo que se debe de tomar en cuenta para escoger la tecnología que se usará.

Con cualquiera de las dos tecnologías se pueden minimizar los tiempos de procesamiento; sin embargo, en la de teclado se requiere de mayor personal, supervisión y procesos de control que en la tecnología de reconocimientos ópticos.

Actualmente existen otras formas de procesamiento, por ejemplo, la utilización en campo de boletas electrónicas que se ingresan en Palms y sólo se centralizan en computadoras que continúan con el resto de procesamiento. Este tipo de procesamiento es muy costoso debido a que se requerirían demasiadas Palms para el levantamiento de un censo; sin embargo, se puede pensar en la utilización para encuestas o métodos de venta para otras instituciones después que se han utilizado las Palms en un censo.

4 MANEJO DE DATOS Y PRESENTACIÓN DE RESULTADOS

4.1 Sistemas de gestión de datos

El procesamiento de datos requiere más que sólo capturar la información y convertirla en datos en computador, dar consistencia a los datos y tabularlos para su publicación final. Se requiere un control sobre los procesos implicados en tiempo real; es decir, medir, analizar y controlar sucesos de los procesos; para ello se requiere un componente de adquisición de los datos en cada proceso, un componente de análisis de los datos y un componente de salida para que pueda mantenerse la respuesta en tiempo real.

Para procesar los datos es necesario unificar la unidad de trabajo en la cual se van a procesar los cuestionarios, es decir que se debe identificar claramente cómo se va a agrupar en campo la recolección de dato (unidad de empadronamiento) y si esa unidad, esa agrupación, es la que se utilizará para controlar el procesamiento. Para efectos de ejemplo en esta sección indicaremos que la unidad de medida es una “caja”, la cual contiene identificación del departamento, municipio, sección y sector donde se ha recopilado la información.

A continuación se describen los procesos principales que han de ser controlados por el *software* de gestión de datos.

4.1.1 Cobertura y completitud

Para tener el éxito esperado en un levantamiento de campo, como un censo o encuesta, se requieren informes preliminares que tengan información básica sobre cantidades esperadas de boletas en estudio (pueden ser viviendas, hogares, personas, fincas, unidades económicas, etc.). Ello requiere que previo al levantamiento del censo o las encuestas se realicen actualizaciones cartográficas que determinen valores de control por unidad política-administrativa. Las fuentes de información de control pueden ser propias o dadas por entidades externas que tengan dicha información

La información de control obtenida requiere procesarla en un sistema que contenga las características de estructura similar al que se utilizará en la captura de datos, es decir que se tengan tablas de información al mismo nivel administrativo en el cual se trabajará el resto del procesamiento. Es necesario tener en el sistema algunos datos de control para su uso posterior en la captura; además, se pueden elaborar sistemas de control para la entrada de boletas a las bodegas de almacenamiento con el fin de tener más fuentes de comparación. Esto implica que se deben tener prioridades y más credibilidad sobre algunos datos que sobre otros.

Puede que sea necesario elaborar un sistema parametrizable tanto en sus niveles administrativos de trabajo como en la cantidad de datos fuente que se le ingresaran.

Un ejemplo de cómo sería la información requerida, sería para el desarrollo de un censo de población y habitación:

Departamento	Municipio	Sección	Sector	Hombres	Mujeres	Locales Particulares	Locales Comerciales	Fuente de Información
02	01	001	1	1860	1976	28	36	01
02	01			5470	5800			06

Como se puede ver en la tabla anterior, se pueden tener en una misma tabla a diferente nivel administrativo, información útil que puede servir de control. En la tabla también se ve el campo “fuente de información” que nos indica con la ayuda de otra tabla auxiliar de dónde proviene dicha información.

En resumen se puede decir que lo que se requiere es la recopilación de toda la información local y externa disponible sobre los datos, la cantidad de datos que se estarán estudiando para poder tener controles de cobertura y completitud del dato.

4.1.2 Proceso de conversión del dato

Cuando ya se cuentan con los cuestionarios suficientes, se debe iniciar con el proceso de conversión del dato. Esto conlleva controles para que no se pierda ninguna información en este proceso. Dentro de los puntos de control que debe incluir este sistema son:

4.1.2.1 Control de lo que ingresa para ser digitado

Se debe llevar un registro exacto de lo que ingresa a procesamiento de datos (se recomienda que sean ingresos por unidades administrativas completas), esto es la cantidad de cuestionarios por caja, por tipo de cuestionario (si hubiese más de un cuestionario), fecha de ingreso, hora de ingreso y otros controles que puedan ser útiles para medir el tiempo de proceso.

4.1.2.2 Control de asignación de trabajo

Dependiendo la estructura organizacional que se escoja para procesar los datos, es necesario que se tengan controles de asignación de trabajo al personal. Este módulo debe responder a preguntas como ¿quién tiene la caja?, ¿qué caja está pendiente de asignar?, ¿en qué proceso se encuentra cierta caja?, ¿ya se devolvió la caja terminada?, y otras preguntas que puedan surgir. Además es en este módulo donde se deben realizar las comparaciones y enlaces con el Módulo de completitud y cobertura, contando los datos digitados y comparándolos con las fuentes de información que se tengan a disposición.

4.1.2.3 Control de calidad del dato

Independiente del sistema de captura que se utilice, es necesario desarrollar controles de calidad del dato. Esto es posible al reingresar muestras de la información digitada y compararla. Como se indicó en el capítulo III, si es ingreso por teclado la mayor parte de *software* existente para ello cuenta con modos de verificación.

Es necesario que en este módulo se controle que no sea la misma persona que digitó la que verifica y llevar historial de porcentajes de error para determinar si es necesario la redigitación completa.

4.1.2.3 Control de rendimientos de trabajo y producción

Este módulo debe aprovechar los recursos que tienen los módulos de captura de datos sobre rendimientos y hacer la explotación de los mismos en reportes. Algunos ejemplos de los reportes a tratar en este módulo son:

- Rendimiento promedio por persona
- Rendimiento promedio por grupos de trabajo

- Reporte de avance de procesamiento
- Reporte de reingresos (lo que está mal digitado)
- Comparativos contra metas
- Producción diaria

Dentro de los sistemas de gestión de datos, éste es de los que mayor cuidado y trabajo requieren ya que de ello depende que el procesamiento de datos se realice en el tiempo estipulado. Normalmente la cantidad de personal y datos muy grande por lo que requiere de buenos sistemas para controlar al personal y la información.

4.1.3 Sistema de apoyo

Además de los sistemas que se requieren para el procesamiento, se debe tener presente que hay que dar apoyo a otras actividades dentro del trabajo de realización de un censo o encuestas.

Algunos de los sistemas más comunes de apoyo son:

4.1.3.1 Recursos humanos

Control del personal que labora en las diferentes etapas del censo y encuesta. Debe incluir facilidad de pago para personal temporal y permanente.

4.1.3.2 Control de activos fijos

En un proceso tan grande como los censos o elaborar encuestas grandes se requiere tener control de todo el equipo (computadoras, automóviles, escritorios, etc.) que se utilizan para su desarrollo. En este tipo de sistemas se debe de llevar registro exacto del mobiliario y equipo con fechas de adquisición, proveedores, garantías y responsables del equipo para facilitar cualquier gestión.

4.1.3.3 Contabilidad y bancos

Si el control de los ingresos y gastos se llevara a cabo internamente, se debe de pensar en controles de la contabilidad y elaboración de cheques para agilizar todos los movimientos administrativos que se lleven a cabo dentro de los proyectos.

4.1.3.4 Apoyo a ejecución de campo

Dentro del proceso de elaboración de un censo o una encuesta, uno de los factores más complicados es el de levantar la información en campo. Ello requiere buenos controles para saber qué persona está trabajando cierto sector administrativo de levantamiento de datos.

Normalmente esto se hace descentralizado por lo que los sistemas de apoyo en este tipo de trabajo se deben realizar meticulosamente y planificar con tiempo la forma en la cual se centralizarán los datos.

4.1.3.5 Apoyo a la cartografía

Como se explicó en el capítulo I, una de las actividades centrales para elaborar un censo o encuesta es la de actualizar la cartografía de campo. Posiblemente esto genere algunos formularios que hay que procesar lo cual recae en el departamento de procesamiento de datos. También en el caso de un censo . esto genera cargas de trabajo administrativas; es decir, dividir los departamentos y municipios en áreas de trabajo; esto se utiliza entre censo y censo como cargas de trabajo para encuestas por lo que se debe generar la información de un marco de sectores.

En estos sistemas de apoyo se pueden incluir otros, dependiendo de la necesidad del trabajo; sin embargo, es importante que se evalúe el hecho de que ya existen controles de este tipo en el mercado a precios razonables, lo cual pueda ser mejor obtenerlos que desarrollarlos internamente. Esto redundaría en que el personal de procesamiento de datos se centrara principalmente en los módulos que tienen que ver con el procesamiento mismo.

4.2 Presentación de resultados

4.2.1 Objetivos

Los principales objetivos de censos y encuestas son recolectar, compilar y hacer conocer al público información útil y aprovechable.

Se puede decir que el censo está completo solamente cuando esta información ha sido puesta a disposición de aquéllos que la usarán para propósitos prácticos de toma de decisión, investigación, planificación y administración rutinaria de gobierno. La preparación de los datos censales para el propósito de distribución cumple así con el objetivo último del censo. Todo esfuerzo del procesamiento de datos en términos de insumo humano, tecnológico y financiero está representado por los datos que se publican. Por lo tanto, las publicaciones deben ser cuidadosamente planificadas para maximizar su utilidad y cumplir con los requerimientos de los diferentes usuarios utilizando el menor tiempo posible después de procesados los datos.

4.2.2 Planificación

Mientras la utilidad de la información que va a ser recolectada se analiza en las primeras etapas de planificación de la investigación de campo (cuando se elaboran los cuestionarios), la practicabilidad o facilidad de uso de la información se determina en gran parte al final de la investigación, durante la fase de publicación.

Es en este momento cuando se debe prestar especial atención y se deben dar todos los pasos posibles para asegurar precisión, presentación clara y accesibilidad de los resultados de datos.

El principio que guía el trabajo hecho en esta etapa, como en otras etapas del censo y encuesta, es que todas las decisiones sobre detalles técnicos de procesamiento y elaboración de las publicaciones debe concentrarse en el uso que se dará a la información, es decir, el énfasis debe ponerse principalmente en las necesidades de los usuarios del dato.

Para planificar cómo y en qué se van a entregar los datos, es necesario revisar si se tendrán ayudas analíticas tales como mapas temáticos, gráficas, estadísticas derivadas (por ejemplo, distribuciones porcentuales y medianas) y texto descriptivo. Normalmente todos los datos que se recolectan y tabulan en los censos y encuestas se pueden poner a disposición del público (sin dar identificación de los hogares donde fue tomada la información para mantener la confidencialidad). La mayoría de los datos se emiten en la forma de informes impresos, CD con tabulados y mapas, servicio de información en la web, etc.

Además se debe de determinar el detalle geográfico que se dará a los datos publicados. El número de áreas geográficas sobre las cuales se entregarán datos censales tiene una influencia obvia en el alcance del programa de publicación. La forma de presentar el detalle geográfico también tiene impacto. Lo ideal sería publicar series distintas de publicaciones que contuvieran categorías de información algo sobrepuestas. Podrían presentar publicaciones orientadas a áreas temáticas y otra a áreas geográficas por tema. La cantidad de información publicada para los diversos niveles geográficos merece importante consideración. Diez páginas de cuadros para el total del país no es significativo, pero 10 páginas para cada una de las 7,282 distintas áreas administrativas del país haría una diferencia considerable.

Los cuadros estadísticos son los principales instrumentos usados para organizar y dar forma a la parte substancial de las investigaciones censales y de encuestas, esto es, la información que ha sido recolectada en el empadronamiento y procesada por computadoras.

En términos físicos, la mayor parte del informe esta constituida por cuadros estadísticos.

Su preparación consume la mayor cantidad de tiempo y de atención y se requiere que éstos respondan a especificaciones técnicas exigentes. Las pautas escritas para la construcción de los cuadros deben prepararse con mucho detalle y hacerse circular entre todos los involucrados en el procesamiento de datos. Con la ayuda de computadoras y programas especializados, el tiempo requerido para tabulación de datos se ha reducido considerablemente. Con una adecuada planificación, el *software* puede producir cuadros definitivos listos para distribuir a los usuarios.

Cuando se desarrollen las publicaciones es necesario tomar en cuenta lo siguiente:

El formato: el formato de las publicaciones debe observar normas de buena presentación de los datos estadísticos en forma de cuadros o gráficos. Debe ser lo bastante general y lo bastante específico como para responder a las necesidades de todo tipo de usuario, a las de aquéllos con preparación estadística formal y a las de los que no tengan ninguna. El informe debe contener cualquier soporte analítico que se considere necesario para hacer más comprensibles los resultados del censo y asegurar que se usen correctamente. De este modo, el formato no es un ordenamiento opaco y poco interesante de los cuadros estadísticos, gráficos, texto e índices. Tiene una importante función llevar toda la operación censal al punto donde pueda conseguirse el objetivo del informe, que es hacer que los resultados se usen al máximo.

Presentaciones claras: para alcanzar este propósito, el informe debe presentar claramente los datos. Los datos deben ser ordenados en forma comprensible y estructurada. Es más, el formato debe ser consistente, sencillo y ordenado.

Consistencia en tabulados: es necesario ceñirse estrictamente a las reglas de presentación de datos para evitar confusión, y hacer posibles comparaciones válidas de información presentada en diferentes partes del mismo informe o en otros informes censales. Deben escribirse las reglas concernientes a estructura de cuadros, elaboración de mapas y construcción de gráficas y otras ayudas visuales. También deben darse ejemplos claros de cómo se van a aplicar las reglas en diversas situaciones. Las reglas deben cubrir todos los detalles: los títulos que requieren redacción uniforme y un orden prescrito para describir el contenido, el uso y elaboración de notas al pie y de notas de encabezamientos, uso consistente de mayúsculas, abreviaturas, puntuación, sangría y uso de espacios verticales y horizontales; uso de símbolos para denotar celdas no aplicables, etc.

Sencillez: un diseño formal puede reducir la posibilidad de distorsión o de mala interpretación de los datos simplificando el contenido de cada cuadro, mapa o gráfico y usando un sistema claro y uniforme de anotación estadística. El contenido de cada cuadro, mapa o gráfico debe limitarse a expresar una idea, aunque pueda ser una idea compleja. Al clasificar en forma cruzada los datos básicos con otras características, se debe tener cuidado de introducir en el cuadro sólo aquellos elementos de información que contribuyen a aclarar el punto principal que se está tratando. Reducir los datos de un cuadro a una forma visual tal como un gráfico o un mapa puede demostrar que el cuadro está tratando de decir demasiado.

Si se requiere más de un gráfico para expresar la idea, o si la información llevada al papel se ve demasiado confusa, el cuadro mismo puede ser confuso.

Orden: la información detallada se capta más fácilmente si está organizada en categorías significativas.

Las publicaciones deben diseñarse para mostrar una progresión lógica de información, empezando con distribuciones simples y temas generales y terminando con los temas más específicos y detallados. Esta progresión no sólo se aplica al detalle del tema, sino también al detalle del área geográfica. En términos de presentación mecánica, es habitual mostrar primero la información sobre el país, en segundo lugar la información sobre las divisiones administrativas mayores, enseguida la información sobre áreas administrativas menores y después la información sobre localidades, llegando hasta las áreas más pequeñas para las cuales se van a publicar datos.

4.2.3 Software de presentación de resultados REDATAM+ G4

4.2.3.1 Qué es Redatam+G4

Redatam es el acrónimo de **RE**cuperación de **DAT**os para **Á**reas pequeñas por **M**icrocomputador. **Redatam+G4** (abreviado **R+G4**), la última versión, es la cuarta generación del *software*, que puede ser usada en español, inglés o portugués, con Microsoft Windows 95, 98, NT4 o 2000, en cualquier microcomputador compatible con IBM (*winR+* v.1.2 es la versión previa, que opera también con Windows 3.1).

El programa utiliza una base de datos jerárquica comprimida, que se puede crear en *R+G4* y que contiene micro datos y/o información agregada con millones de registros de personas, viviendas, manzanas de ciudades o cualquier división administrativa de un país. Esos datos pueden provenir de cualquier combinación de censos, encuestas u otras fuentes.

Se puede procesar una base de datos en asociación con bases de datos externas de formatos comunes, como dBASE. Bases de datos jerárquicas muy grandes y no geográficas también han sido creadas en tales campos, como comercio internacional.

Es posible definir, a partir de una base de datos, cualquier área geográfica de interés (desde manzanas de una ciudad) o combinaciones de esas áreas, crear nuevas variables, obtener varios tipos de tabulados muy rápidamente y exportar salidas a otros *softwares*.

Los datos de diferentes niveles geográficos pueden ser combinados jerárquicamente para crear variables agregadas, y los resultados pueden desplegarse en mapas desde Redatam o transferirse a un Sistema de Información Geográfica (SIG). Se usan versiones de Redatam en la mayoría de los países de América Latina y el Caribe, en varios de África y Asia y en muchos países desarrollados; alrededor de 1,600 usuarios e instituciones de 80 naciones son usuarios registrados. La primera versión de Redatam para DOS se publicó oficialmente en 1987. Todas las versiones de Redatam han sido desarrolladas y mantenidas por el Centro Latinoamericano y Caribeño de Demografía (CELADE), que es la División de Población de la CEPAL.

El desarrollo de *Redatam+G4* y otros miembros de la familia REDATAM ha recibido ayuda financiera de CIID, ACIDI, FNUAP, BID y del presupuesto regular de las Naciones Unidas.

4.2.3.2 Qué contiene Redatam+G4

Redatam+G4 (abreviado *R+G4*), la cuarta y más nueva generación del programa Redatam -introducido en 2001-, cuenta en la actualidad con los siguientes miembros, que permiten crear, administrar, procesar y diseminar grandes bases jerárquicas de datos censales y de otro tipo

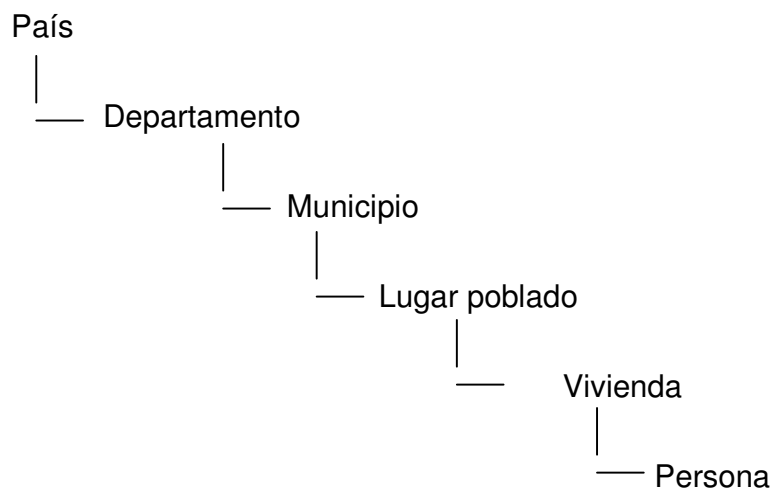
:

R+Process procesa datos provenientes de áreas geográficas seleccionadas por el usuario -para cuyo efecto cuenta con poderosos programas escritos en el lenguaje de comandos Redatam- o procesa con el restringido asistente -parecido al *Wizard* de algunos programas comerciales- que no necesita ninguna habilidad en programación y tampoco conocimiento alguno del lenguaje Redatam.

Este módulo incluye, además, facilidades para expandir, agregar, conectar, descargar y realizar otras acciones, con las actuales bases de datos *R+G4*.

R+Create permite crear bases de datos jerárquicas en un formato propio de *R+G4* a partir de archivos en formato ASCII o xBase. Los archivos ASCII pueden tener sus registros (*layouts*) definidos en IMPS, CPro, ISSA o CHILLAN (formato interno de R+G4).

Para crear una base de datos, podemos considerar un ejemplo de base de datos con una sola jerarquía geográfica, como la que se muestra en la figura siguiente:



Cada uno de los niveles, como país, se denomina una entidad. En cada entidad país se encuentra la entidad departamento, los que a su vez tienen lugares poblados, viviendas y personas. Una base Redatam puede tener más de una rama jerárquica, y éstas a su vez pueden tener sub-ramas. Los conceptos y elementos de entidad son los mismos.

La información sobre los elementos individuales de una entidad se guardan en variables, por ejemplo, el número de kilómetros cuadrados de un departamento o el sexo de una persona. De tal forma, se puede ver una entidad como si fuera matriz, en donde las filas son los elementos (casos) y las columnas son las variables.

Estructura lógica: una base de datos Redatam está definida lógicamente en tres partes, cada una con su función específica:

a) Diccionarios (.dic): toda la información referente a las entidades y las variables se guardan en el diccionario, el cual tiene los meta datos de la base que es información sobre la información. El diccionario de la base de datos permite al usuario visualizar los datos en términos de variables y de registros, alejándolo de los detalles físicos de almacenamiento y manipulación de datos. El diccionario contiene una lista de variables de una base para cada nivel de la jerarquía así como los códigos (valores) para cada variable, con una descripción del significado de cada código. El diccionario para una determinada base generalmente es definido una sola vez. Sin embargo, Redatam presenta opciones para modificar y editar el diccionario y así adecuarlo a los requisitos específicos del usuario. Redatam también posee funciones para usar diccionarios y bases escritos en IMPS y CSPro.

b) Archivos de punteros (índices .ptr): estos archivos conectan los elementos de las entidades con sus entidades inferiores. Cada entidad tiene un archivo de índices con elementos que “apuntan” desde los elementos de la entidad superior hasta los elementos de la entidad misma.

c) Archivos de datos (.rdf): el sistema Redatam guarda cada variable de una entidad en su propio archivo, con un “registro” por cada elemento de la entidad; cada variable es un vector de los datos.

Tales archivos se conocen como archivos transpuestos, para distinguirlos de la forma más común de organización de los datos estadísticos, donde todas las variables de cada elemento se guardan en registros individuales, de manera que un solo archivo contiene todas las variables para todos los elementos.

En la estructura del archivo transpuesto existe un archivo para sexo, otro para tipo de vivienda, etc.

Entre otras ventajas, el procesamiento del sistema es eficiente, ya que permite que el sistema lea solamente las variables específicamente pertinentes a un determinado proceso.

Esquema conceptual: por lo menos cuenta con tres grandes conceptos:

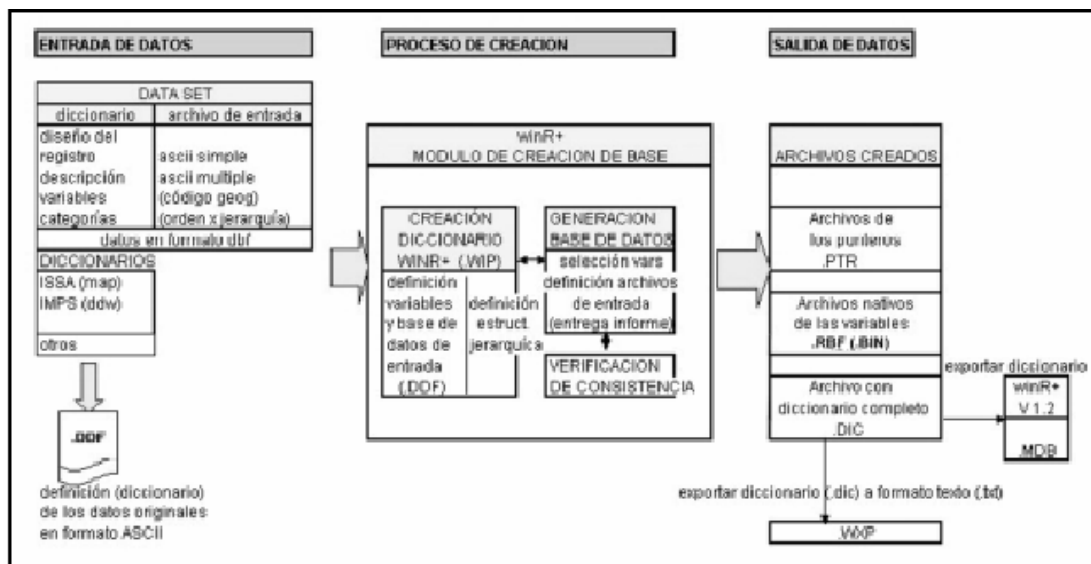
a) Entradas de datos: los datos pueden estar en ASCII con un solo tipo de registro o con dos o más formatos, por Ejemplo, registros para hogares y para personas. Los registros pueden ordenarse por códigos geográficos, para ordenar la jerarquía geográfica; por ejemplo, personas en hogares, manzanas, municipios, departamentos, etc. Los datos también pueden estar en archivos .dbf

b) Información del diccionario: si su diccionario está en IMPS, puede convertir los metadatos directamente en un diccionario Redatam. En caso contrario, se necesita un “libro de códigos” que defina el diseño de los registros, la descripción de variables y los códigos. En todos los casos es muy importante poder relacionar en un mapa los códigos geográficos y las áreas para desplegar las relaciones espaciales de los resultados.

c) **Creación de la base de datos:** el primer paso es convertir los metadatos en un archivo .ddf, que luego se usa con la definición de las jerarquías de la base de datos y de las entidades seleccionadas, etc. Se comprueba la consistencia de los datos y metadatos y se crean tres conjuntos de archivos: índices (.prt), los datos en sí, con un archivo para cada variable (.rdf); y diccionario (.dic).

Los módulos de creación y administración permiten exportar el diccionario como un archivo de texto para impresión y otros fines.

Figura 18. Sistematización de Redatam+G4



R+xPlan permite construir aplicaciones destinadas a los usuarios finales que les permitan acceder, sin ningún conocimiento de Redatam, a microdatos protegidos con el fin de obtener indicadores que con frecuencia implican cálculos complejos.

Figura 19. R+xplan permite la creación de aplicaciones de manejo sencillo

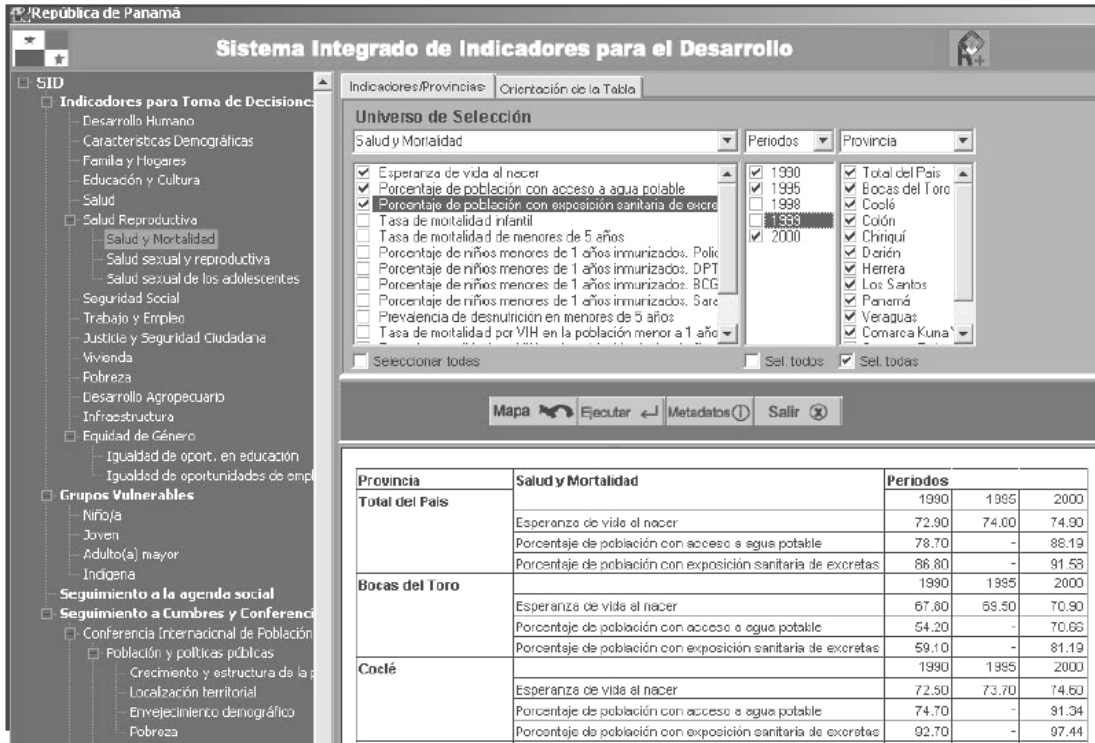
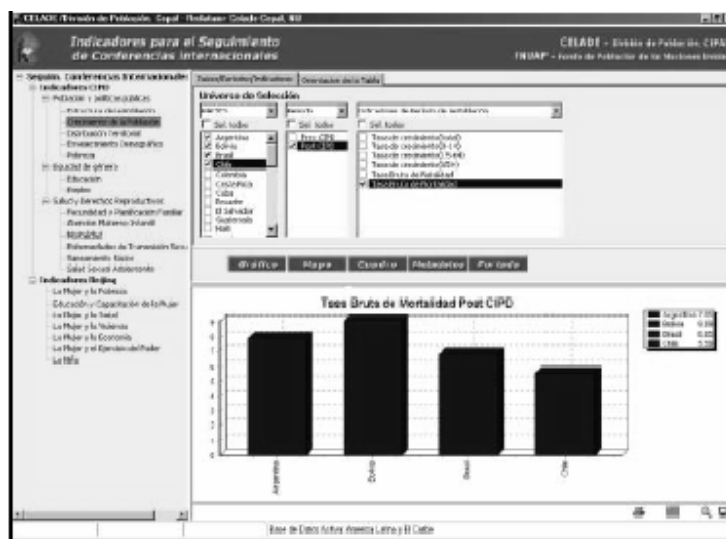


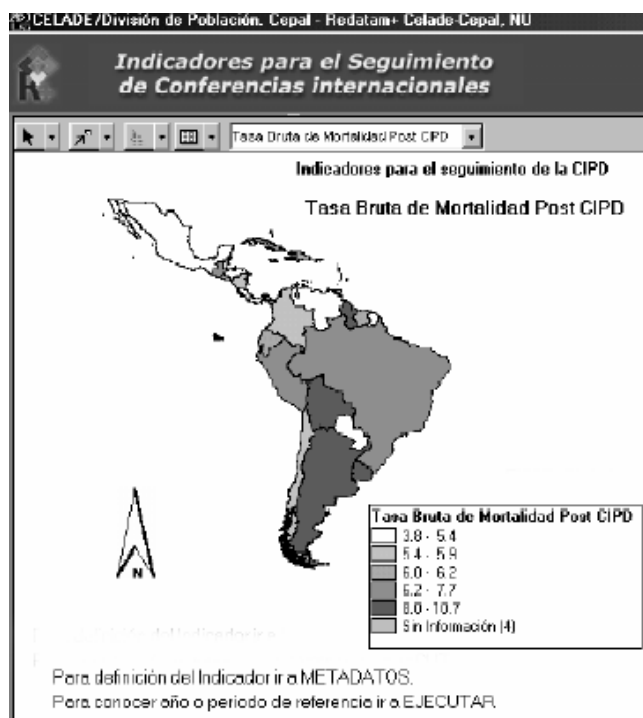
Figura 20. R+xplan permite la obtención de indicadores que con frecuencia implican cálculos complejos de una forma accesible y sencilla



R+WebServer, el más reciente (2002) integrante de la familia, puede solicitar información en línea de una base de datos *R+G4* a través de Intranet o extranet de internet. Al igual que en el caso de los otros miembros de la familia REDATAM, el *R+WebServer* se funda en la maquinaria estadística REDATAM, que posee una alta eficiencia en un ambiente en línea.

Sistemas Redatam+GIS para la planificación: a mediados de 1990, y en el marco del proyecto conjunto CELADE-Universidad de Waterloo (Canadá), financiado por el IDRC de Canadá a mediados del decenio de 1990, la Universidad creó un conjunto de herramientas para apoyar la toma de decisiones en el plano espacial, que utiliza un Sistema de Información Geográfica (SIG) propio.

Figura 21. Redatam ha creado herramientas para la toma de decisiones en el plano espacial de información geográfica



Los tres primeros miembros de la familia Redatam forman un solo conjunto de módulos del programa Redatam, y éste puede descargarse, sin costo alguno, del sitio Web del CELADE <http://www.eclac.cl/celade/redatam> . El servidor *R+WebServer* opera con un servidor Windows.

Actualmente existen varios *software* en el mercado que son funcionales para elaborar publicaciones; sin embargo Redatam es un *software* que se está utilizando en las diferentes instituciones estadísticas de los países como estándar para presentar los datos de censos y encuestas. Las publicaciones con los datos se deben realizar con calidad y facilidad de manejo en un corto tiempo y hacerlas disponibles para el público en general. En esto contribuye Redatam que es una herramienta para distribución y presentación de resultados que es funcional, fácil de utilizar y escalable para presentar todo tipo de resultados.

CONCLUSIONES

1. El procesamiento de datos de los censos y encuestas es uno de los procesos más grandes que existen dentro de un país y también uno de los más importantes para tomar decisiones tanto de tipo empresarial como gubernamental, por lo que requiere personal capacitado en el campo de las ciencias de la computación para su desarrollo.
2. Existen actividades antes del procesamiento que determinan cómo elaborar el procesamiento de datos. Todas las actividades deben ser conocidas y apoyadas por el personal de procesamiento de datos, ya que no puede existir independencia total dentro de los procesos.
3. Una de los procesos más grandes y complejos es el de convertir el dato que viene de campo en dato consistente y listo para trabajar como bases de datos; aunque se pueden desarrollar programas a la medida, se debe de tener en cuenta que existen programas ya elaborados y probados por instituciones de renombre para censos y encuestas, los cuales son de calidad y cumplen con los requerimientos necesarios. Por lo tanto es mejor utilizar estos programas uniéndolos con programas de control que elaborar todos los programas.

4. Actualmente existe tecnología de punta de reconocimiento óptico que facilita y agiliza los tiempos de conversión del dato; es necesario evaluar los elementos externos (nivel de educación de empadronadores, ambiente físico del lugar de procesamiento, etc.) para decidir si se utiliza este tipo de tecnología.
5. Si se utiliza el método por teclado, con los sistemas actuales es posible entregar resultados oportunamente. Por ejemplo, en los censos anteriores al de 2002, la información requería de un promedio de 3 a 4 años para estar lista y procesada en su totalidad. Actualmente sólo se requiere menos de 1 año para tener la información lista para los usuarios.
6. Todos los puntos del procesamiento son importantes, pero no tendría ningún sentido efectuar grandes esfuerzos por procesar los datos si no se divulgaran. Para ello requiere una buena planificación del tipo de publicación que se elaborará y utilizar programas que den las facilidades para publicar los datos.
7. Se puede pensar en distribuir bases de datos con la información, siempre y cuando existan talleres de uso de las mismas.

RECOMENDACIONES

1. Debido a la importancia que tiene, sobre la toma de decisiones el procesamiento de una encuesta y/o censo, se recomienda incluir dentro del área de *software* de la carrera, dentro de la administración de proyectos, información sobre cómo manejar proyectos de procesamiento masivo.
2. Promover dentro de los estudiantes de ingeniería de sistemas el conocimiento y uso de las herramientas existentes de procesamiento de encuestas y censos.
3. Pedir apoyo a las entidades nacionales e internacionales para elaborar talleres sobre el procesamiento, publicación y centralización de datos estadísticos del país.
4. Que los estudiantes de Ingeniería en Ciencias y Sistemas puedan apoyar la creación (de no existir) y fortalecimiento de las unidades estadísticas dentro de la facultad de ingeniería y la Universidad de San Carlos.

BIBLIOGRAFÍA

- Informe metodológico, X Censo de Población y V de Habitación. República de Guatemala. 1998.
- Manual de CPro. Buró de Censos de Estados Unidos. 2004.
- Manual de Encuesta Nacional de Gastos e Ingresos Familiares –ENIGFAM-. República de Guatemala. 1999.
- Manual de IMPS. Buró de Censos de Estados Unidos. 2004.
- Manual deL Módulo de Creación de Bases de Datos en REDATAM+G4. CELADE/CEPAL. 2004.
- Manual deL Módulo de R+xPLan en REDATAM+G4. CELADE/CEPAL. 2004.
- Manual deL Módulo de Procesamiento de Datos en REDATAM+G4. CELADE/CEPAL. 2004.
- U.S Bureau of the Census, Popstan: Un Estudio de Caso para los Censos de Población y Habitación de 1980, Parte A, Consideraciones Programáticas, Documento de Adiestramiento Estadístico, ISP-TR-4^a, Washington, D.C., 1980.
- Vijverberg, Wim. Measuring Income from family Enterprises with Household Surveys. Living Standards Measurement Study. World Bank, Washington, D.C. 1991.