



PROCESAMIENTO

Métodos

Enero 2017

Etapas

1

- Levantamiento Precenso (Sectorización del país y Cantidad de Cuestionarios por Portafolios)

2

- Diseño fino de los Cuestionarios Censales.

3

- Impresión de Cuestionarios Censales.

4

- Mecanizado de Portafolios y Cajas.

5

- Distribución de Cajas a bodegas comunales.

6

- Levantamiento Censo 2017

7

- Logística reversa del material censal

8

- Procesamiento Primario

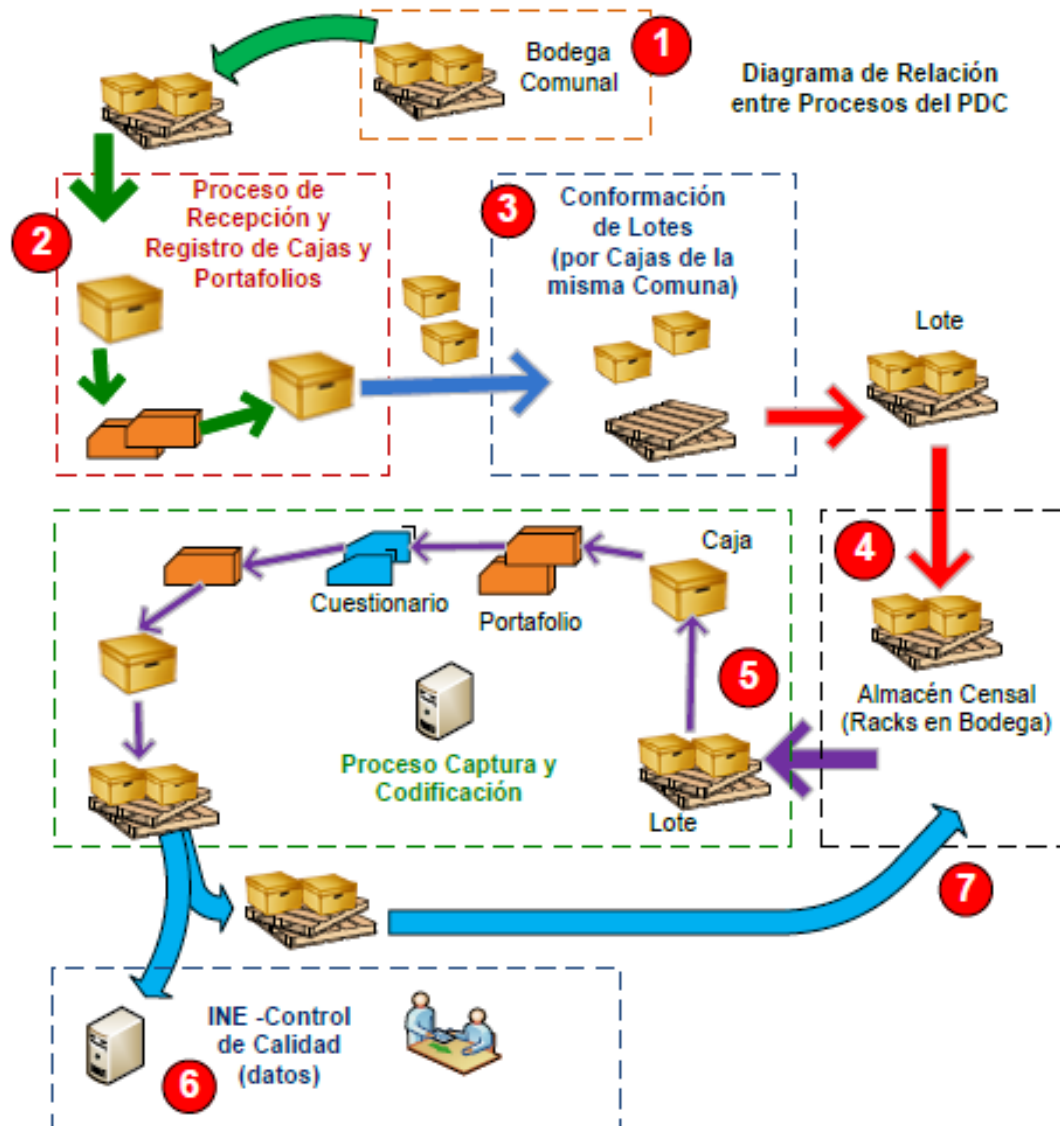
9

- Procesamiento Secundario

10

- Diseminación de los resultados Censales

Procesamiento de datos (Procesos)



Control de
completitud
permanente

Procesamiento de datos (BD)

Procesamiento Primario

Procesamiento secundario

Empresa adjudicada



Reconocimiento
óptico (BD Pura)

Correcciones
in situ
Verificaciones
Codificaciones

BD entregada
por empresa

Fase previa
Consistencia
estructural
Estructura BD

Validaciones y
corrección de
inconsistencias

Etapas

1

- Levantamiento Precenso (Sectorización del país y Cantidad de Cuestionarios por Portafolios)

2

- Diseño fino de los Cuestionarios Censales.

3

- Impresión de Cuestionarios Censales.

4

- Mecanizado de Portafolios y Cajas.

5

- Distribución de Cajas a bodegas comunales.

6

- Levantamiento Censo 2017

7

- Logística reversa del material censal

8

- Procesamiento Primario

9

- Procesamiento Secundario

10

- Diseminación de los resultados Censales

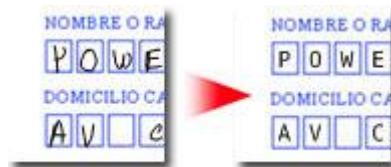
Procesamiento Primario

OBJETIVO:

Obtener una base de datos estructurada, producto del reconocimiento del contenido de escritura y marcas de los cuestionarios ocupados en el levantamiento censal para proceder, en la etapa posterior, a validar y corregir inconsistencias.



- Captura digital de cuestionarios censales



- Reconocimiento óptico de marcas y caracteres
- Codificación automática y asistida
- Consistencia estructural



- Fijaciones de variables base
- Estructura por Entidad: Vivienda, Hogar y Personas

Procesamiento Primario

Conformación lotes de trabajo



Preparación del Material



Digitalización



Reconocimiento óptico



Verificación de textos



Verificación de marcas y números



Codificación Automática y Asistida



Control de calidad



Control de consistencia estructural



Generación de formato de salida



Entrega de BD e imágenes a INE



Fase Previa al Procesamiento
Secundario

Conformación de lotes y preparación



1. Retiro de pallet con las correspondientes cajas desde el archivo Censal
2. Se retiran las cajas del pallet para su preparación
3. Se abren las cajas con los Portafolios en su Interior
4. Se revisará visualmente del contenido de la caja, en caso de que vengan documentos en un estado no compatible con el proceso, se enviará a un proceso de manejo de excepciones
5. Se procederá a hacer un conteo de Portafolios de cada Caja, registrando esto en el sistema de archivo
6. Se emitirá la "Hoja Conductora" de acuerdo al real contenido de la caja
7. En este punto se envía la caja con la "Hoja Conductora" a preparación, orden y completitud inicial
8. Guillotinado y preparación de lotes en hojas individuales para comenzar la digitalización.

Digitalización y Reconocimiento óptico



1. Escaneo de imágenes de los cuestionarios censales.
2. Almacenamiento de imágenes en BN y COLOR para optimizar la verificación.
3. Reconocimiento óptico de marcas y caracteres (OCR/ICR/OMR)
4. Control de cajas y portafolios.
5. Control de completitud y duplicidades.
6. Reportes de gestión.

VERIFICACIÓN

1. Verificación interactiva de textos, marcas y números de un operador en pantalla
2. Revisión de imagen y base de datos de aquellos casos en que el reconocimiento automático no logra el nivel de confianza deseado.

CODIFICACIÓN AUTOMÁTICA Y ASISTIDA

1. Codificación de variables abiertas, a través de algoritmos automatizados y de verificación en pantalla.
2. Personal INE.

1. Control de calidad in situ del reconocimiento óptico.
2. Control de integridad para completitud
3. Control de consistencia estructural

Formato de salida y entregas

1. Los datos son entregados por el contratista en formato de respaldo de BD MS SQL Server. A razón de un registro por Cuestionario completo, y, donde los campos son exactamente los campos definidos en el proceso de reconocimiento de los cuestionarios.
2. Se adicionan campos de:
 1. Codificación (valor y código normalizado de glosas)
 2. Información del contexto, como es el número de orden de cada cuestionario dentro del Portafolio.
 3. Información de la identificación geográfica al cruzar con la BD de Portafolios.
4. Entregas programadas de BD e imágenes, por lotes al menos una vez por semana.
5. Control de calidad (BD e imágenes) para recepción conforme. Tasas de reconocimiento.

1. Fase previa al proceso de validación y corrección de inconsistencia.
2. Normalización de información estructural.
 - Portafolio
 - Vivienda
 - Hogar
 - Continuación
 - Ocupación
3. BD relacional
 - Vivienda
 - Hogar
 - Personas

Etapas

1

- Levantamiento Precenso (Sectorización del país y Cantidad de Cuestionarios por Portafolios)

2

- Diseño fino de los Cuestionarios Censales.

3

- Impresión de Cuestionarios Censales.

4

- Mecanizado de Portafolios y Cajas.

5

- Distribución de Cajas a bodegas comunales.

6

- Levantamiento Censo 2017

7

- Logística reversa del material censal

8

- Procesamiento Primario

9

- Procesamiento Secundario

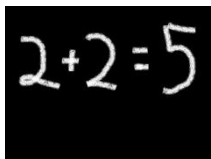
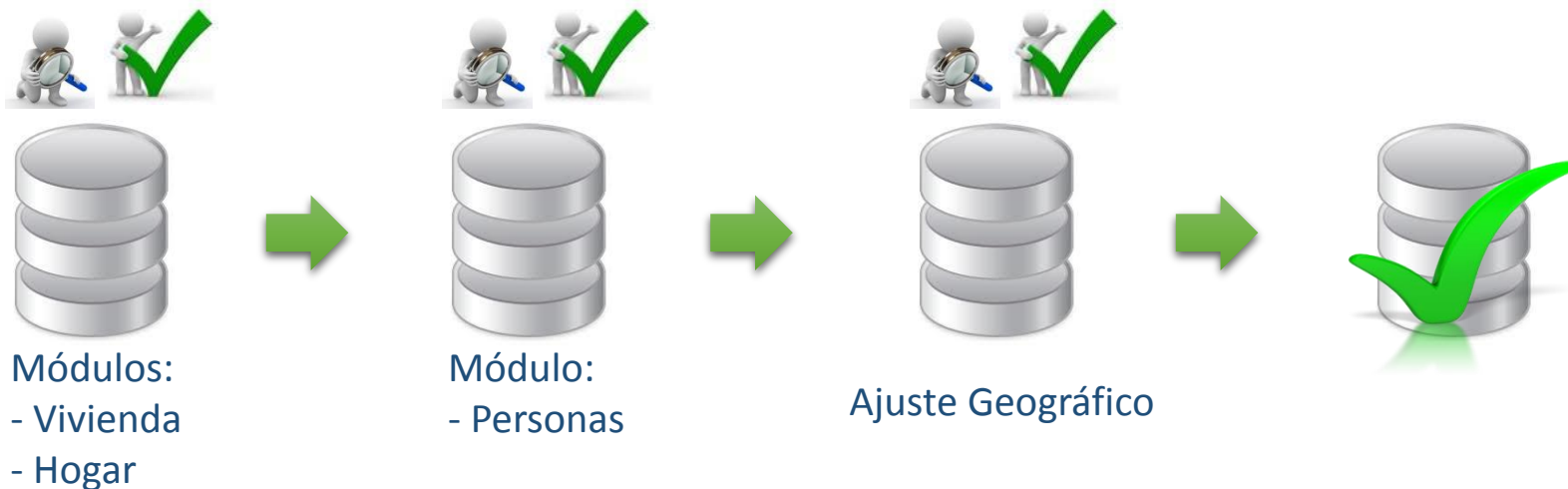
10

- Diseminación de los resultados Censales

Procesamiento Secundario

OBJETIVO:

Obtener información consistente, de la base de datos generada en el procesamiento primario, de viviendas, hogares y personas, para los fines que se estime pertinente, a través de la ejecución de **procedimientos de validación y corrección de inconsistencias** de esta información censal.



Etapas

1

- Levantamiento Precenso (Sectorización del país y Cantidad de Cuestionarios por Portafolios)

2

- Diseño fino de los Cuestionarios Censales.

3

- Impresión de Cuestionarios Censales.

4

- Mecanizado de Portafolios y Cajas.

5

- Distribución de Cajas a bodegas comunales.

6

- Levantamiento Censo 2017

7

- Logística reversa del material censal

8

- Procesamiento Primario

9

- Procesamiento Secundario

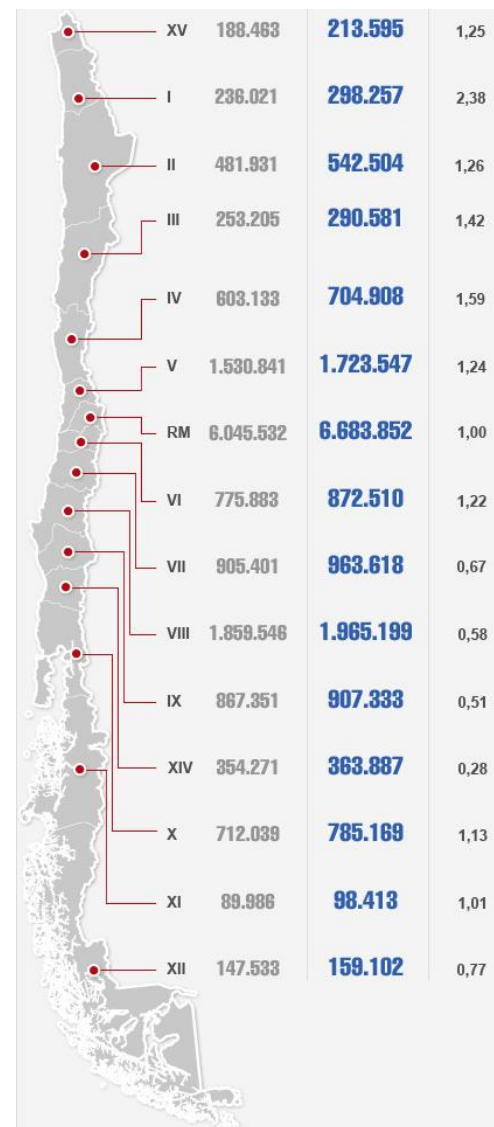
10

- Diseminación de los resultados Censales

Diseminación de Resultados



- BD innominada e indeterminada
- Cartografía consistente con la BD
- Tabulados para publicación análoga y digital
- Redatam
- SPSS u otros formatos



21,57%
de la población es
menor de 15 años

32,8
es la edad promedio
de la población de
Chile

11,8%
de la población de
Chile pertenece a un
pueblo indígena

281.054
de los encuestados
nunca asistió al colegio

19.186
encuestados posee
un doctorado

Línea de tiempo



- Automatización de ejecución
- Herramientas y Lenguajes
- Gestión de versiones de software
- Reportes de diagnóstico

Automatización de ejecución

- Todo el procesamiento está definido en una malla de procesos a ejecutar (Make/Makefile).
- Abarca desde las bases recepcionadas (B0,B1,B2) hasta cada uno de los productos del procesamiento (base definitiva, tabulados primarios, BD indeterminada/innominada)
- La concepción es que siempre es posible reprocesar todo, y regenerar cualquier producto a partir de los datos y configuraciones iniciales
- Los datos y configuraciones iniciales incluye la base de precenso, diccionarios de pcomunas/países, pueblos, etc.

Herramientas y Lenguajes

- Algunas herramientas son “dadas”:
 - La base de datos entregada por contratista es MS SQL Server
 - El procesamiento secundario se hará en CSPRO
 - Tabulados en CSPRO y REDATAM
- Para todo los demás : R
 - Es el lenguaje de scripting del procesamiento (*duct tape*)
 - Usado como lenguaje de propósito general
 - Usado como paquete estadístico
 - Usado para generar reportes (*rmarkdown*)
 - Maneja variados formatos de datos (lee y escribe CSPRO)
 - Fuentes/ejecutables en texto, facilitan SCM
 - Procesa en memoria (ventaja y desventaja)
 - Es gratis, pero puedes pagar si quieres

Gestión de versiones de software

- Programas y configuraciones bajo control de versiones (GIT)
- Sistema integrado de Issue tracker y control de versiones, accesible por web
- Versiones y features coordinadas con entregas

Reportes de diagnóstico

- Cada etapa del procesamiento está acompañada de un reporte de diagnóstico
- El reporte de diagnóstico contiene los resúmenes, selecciones de datos o estadísticas en general que describen el resultado de la ejecución de la etapa.
- El objetivo de los reportes integrados en el procesamiento es el de facilitar el diagnóstico de problemas en el procesamiento, particularmente cuando este sea ejecutado reiteradas veces durante el procesamiento del Censo mismo.
- Cada revisión manual de datos que haya sido necesaria es candidata para incluirse en un reporte de diagnóstico.

Otros mecanismos de diagnóstico

- Las librerías desarrolladas para el procesamiento (R) están acompañadas por un conjunto de *Unit Tests* .
- Los programas usan un esquema de logging flexible al estilo de log4j .



PROCESAMIENTO

Métodos

Enero 2017