CENSO 2017
TODOS CONTAMOS

Tecnología en el procesamiento, validación y consistencia de datos

Noviembre de 2017

CEPAL, Santiago de Chile

CONTENIDO

- Antecedentes Generales
- Etapas del procesamiento de datos censales
- Características del procesamiento de datos censales
- Consideraciones finales

Antecedentes Generales

- Tras la declaración de no oficialidad del levantamiento del 2012, se apeló a la tradicionalidad para el Censo 2017

- Históricamente Chile ha realizado reconocimiento óptico para los Censos de Población y Vivienda
- El proceso completo se licita a proveedores externos con experiencia en proyectos similares
- Rol fiscalizador del INE a través de la ejecución de protocolos operativos que aseguren el cumplimiento del contrato y su flexibilidad

Servicios contratados para el Procesamiento de datos censales

Se incluyen todos los procesos, ya que finalmente la optimización de las tasas de reconocimiento óptico dependen de múltiples factores que pueden afectar desde el diseño, la impresión, el mecanizado, etc., hasta la digitalización de los Cuestionarios Censales

1. Diseño fino de los cuestionarios censales
2. Impresión de cuestionarios y adquisición de materiales
3. Mecanizado de portafolios y cajas
4. Distribución y retiro de cajas desde las bodegas comunales
5. Procesamiento de cuestionarios censales
6. Entrega de bases de datos e imágenes a INE

¿Cómo Procesamos los Cuestionarios?

Por la naturaleza del reconocimiento óptico de los Cuestionarios Censales dividimos el procesamiento en 4 grandes etapas:

PREPARACIÓN	PROCESAMIENTO PRIMARIO	FASE PREVIA	PROCESAMIENTO SECUNDARIO
Definir procedimientos e infraestructura a utilizar.	Obtener BD resultante del reconocimiento óptico de los Cuestionarios Censales	Estructurar la BD para procesamiento secundario.	Obtener una Base de Datos consistente

Clave es la etapa preparatoria e incluir al proveedor el Censo Experimental para simular condiciones reales del procesamiento

Cifras del Procesamiento Primario

1.000+ Pallets con 35.000 Cajas con material Procesable y 15.000 Cajas con material no Procesable

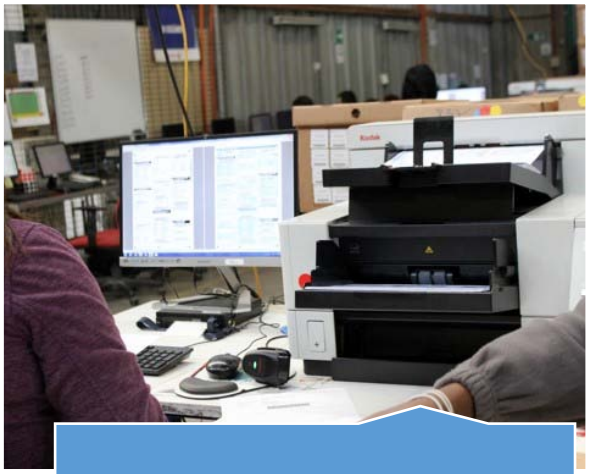
510.000 Portafolios

- 7.100.000 Cuestionarios Procesados
- 70.000.000 de imágenes Color y 70.000.000 Blanco y Negro
- 90 TB de Almacenamiento
- Alrededor de 300 Funcionarios proveedor procesando
- Alrededor de 60 Funcionarios INE en las bodegas del Proveedor
- 12 Escáner de alta productividad operativos y 2 de contingencia
- Más de 20 Entregas de BD e imágenes
- 4 meses de procesamiento con el Proveedor

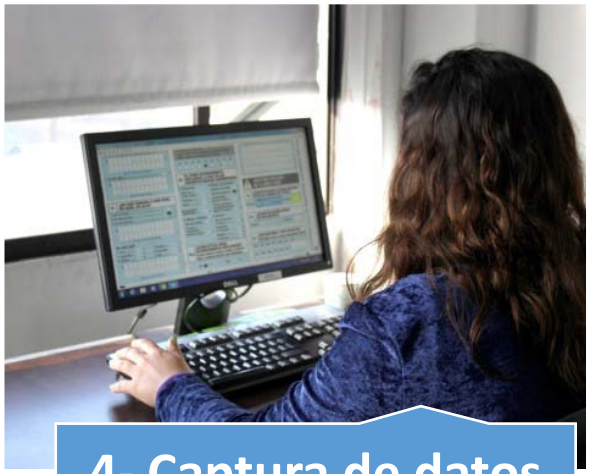
Etapas del Procesamiento Primario

1- Recepción

2- Preparación



3- Digitalización



4- Captura de datos y Verificación

Componentes del Procesamiento Primario
A nivel de plataformas, según etapas

Recepción

Inventario Etiquetas

Preparación Lotes

QA Preparación

Escáner

QA Escáner

Captura de Datos Base0

Verificación Proveedor Base1

Verificación INE Base2

Sistema de Tracking Interno del SITE (Sistema de Trazabilidad)

Módulo Recepción e Inventario

Módulo de Preparación

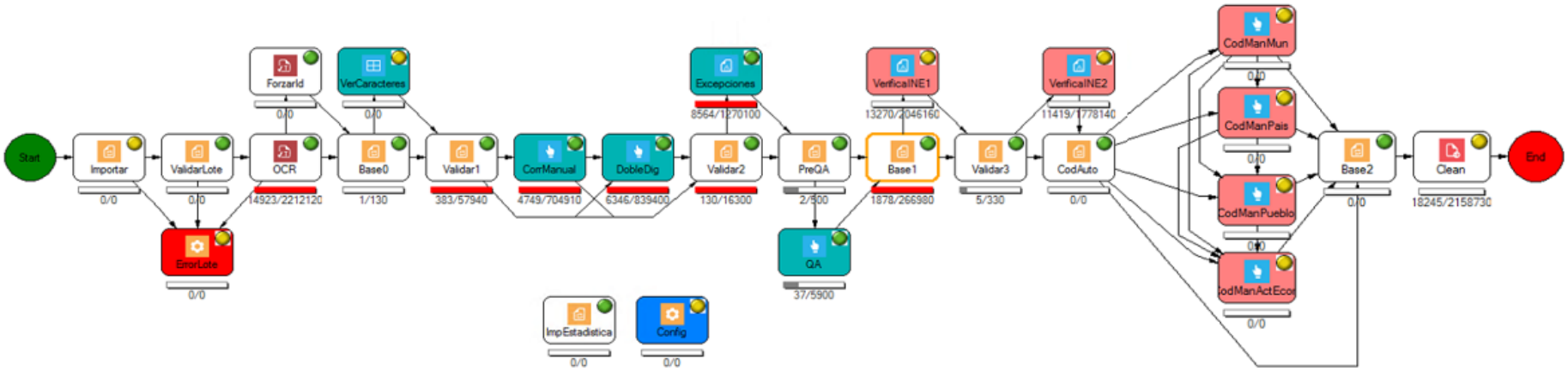
Módulo Scannning

Eflow de TIS

Backup datos e Imágenes

Plataforma con la cual se realiza el reconocimiento óptico: Eflow de TIS

- Sistemas configurable de acuerdo al requerimiento de cada país.
- Funcionamiento modular
- Motores de Reconocimiento OCR /ICR/OMR/BCR
- Validaciones de consistencia
- Visualiza imágenes Color y Blanco y Negro



Factor Clave son los distintos controles durante el Procesamiento Primario

Preparación

- Traspaso de Cuestionarios dañados
- Registro de Cantidades de Cuestionarios.
- QA Preparación
- Excepciones personal INE

Escáner

- 100% Cuestionarios por pantalla
- Mantenciones preventivas
- Validación de Cantidades

QA Escáner

- Revisión del 25%
- Control de propiedades de las imágenes
- Validación de formato y orden
- Captura de código de barra único

Factor Clave son los distintos controles durante el Procesamiento Primario

Eflow y Verificación Proveedor

- Coincidencia de plantillas
- Validaciones de series únicas
- Solicitud de redigitalizar en caso de ser necesario

Verificación INE

- Recorre todas las hojas de los Cuestionarios en la verificación.
- Se revisó el 52% de los Portafolios que equivalen al 54% de las imágenes

Aceptación de Entregas por parte de INE

- Chequeo de Cantidades
- Estructura de la base de datos
- Completitud y comportamiento de variables
- Revisión de una muestra de imágenes versus base de datos

Una vez entregadas las bases de datos al INE, comienza la Fase Previa

ESTRUCTURA VIVIENDAS PARTICULARES	Generar un enlace o vinculación, a través del comportamiento de ciertas variables, entre viviendas-hogares-personas a partir de una base de datos original cuyos registros corresponden a cuestionarios
TRATAMIENTO DE REZAGADOS	Incorporar los cuestionarios de recuperación en local post día del censo y la información capturada vía web, por medio de variables claves como Portafolio/Vivienda/Dirección
UNIFICACIÓN DE BASES DE DATOS	La unificación de Bases de Datos busca generar una base única de datos donde estén los datos de los distintos operativos incorporados (Viviendas Particulares, Viviendas Colectivas, Situación de Calle y Tránsito)

PROCEDIMIENTOS COMPLETAMENTE TRAZABLE

Terminada la Fase Previa, comienza el Procesamiento Secundario

VALIDACIÓN Y CORRECCIÓN DE INCONSISTENCIAS

Todo operativo estadístico presenta distintos tipo de errores según su medio de captura, por lo tanto es necesario conocer las fuentes de error que permitan establecer pautas de revisión y metodologías de corrección.

CODIFICACIÓN DE VARIABLES DE TEXTO

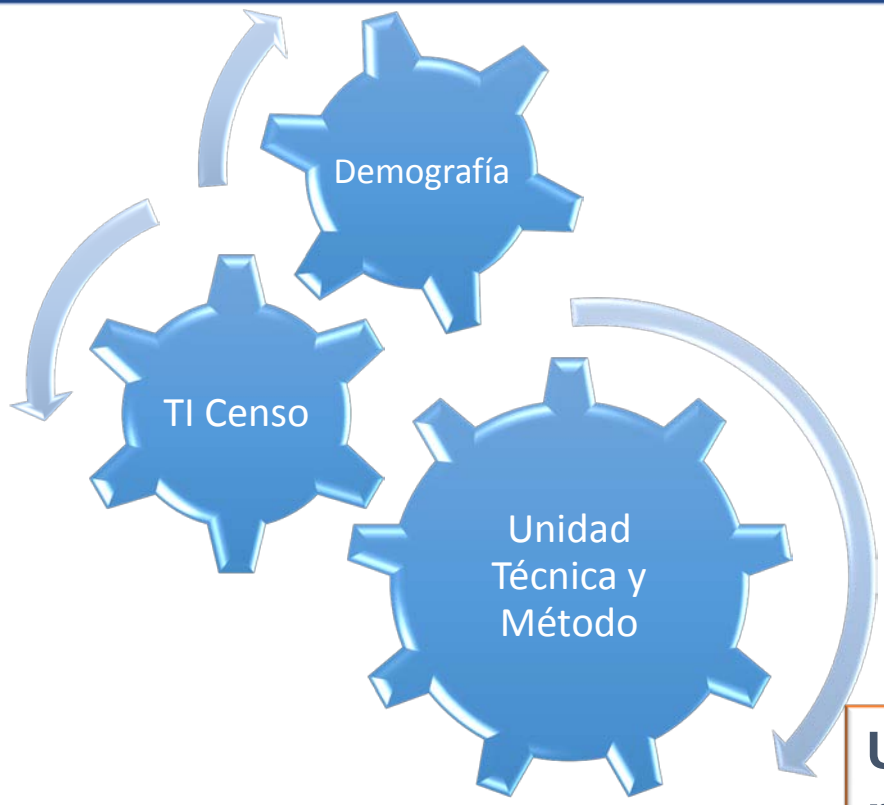
Codificación automática y asistida de la variables de texto de Comunas, Países, Pueblos Originarios y Rama de Actividad Económica. Trabajo interno de INE.

AJUSTE GEOGRÁFICO

Readecuación de los sectores de empadronamiento en el territorio y revisión de los tipos de asentamientos humanos.

PROCEDIMIENTOS COMPLETAMENTE TRAZABLE

Factor Clave: Cohesión de los equipos en el funcionamiento de la sala de Procesamiento censal



La Unidad T y M da los lineamientos a través de:

- **Coordinación técnica del procesamiento**
- **Documentación técnica de los procesos**
- **Revisión de la programaciones de TI**
- **Vinculación con unidades técnicas del INE y externas.**

Utilización de distintos software para revisiones cruzadas de información.

Respuestas oportunas en la programación, revisión y ajustes.

Consideraciones finales

1- Medición del proceso completo a través del aseguramiento de una tasa mínima de reconocimiento óptico.

2- Pruebas escalares de reconocimiento óptico. Incluir al proveedor en el Censo Experimental.

3- Exigir experiencia del Proveedor en procesos similares.

4- Capacidad del proveedor para adaptar software de reconocimiento óptico.

5- Monitoreo y control de cada etapa del proceso y dar flexibilidad al contrato.

6- Conocimiento de las fuentes de error permite incluir controles a lo largo de todo el proceso y focalizar esfuerzos.

Tecnología en el procesamiento, validación y consistencia de datos MUCHAS GRACIAS!!

Noviembre de 2017

CEPAL, Santiago de Chile