

CENSO NACIONAL DE POBLACION Y VIVIENDA 2012

**AREA DE: DIRECCIÓN DE INFORMÁTICA, CARTOGRAFÍA E
INFRAESTRUCTURA ESPACIAL
ETAPA: POST CENSO**



CAPTURA DE DATOS

**MANUAL TÉCNICO DE ESCANEO
DIGITAL**

001 - V.1

2013



POST CENSO

FASE: CAPTURA DE DATOS

"MANUAL TÉCNICO DE ESCANEEO DIGITAL"

Control de Documentación

Código: INE-001-V1 Área Funcional: Dirección de Informática, Cartografía e Infraestructura Espacial

Prohibida su reproducción sin previa autorización. El incumplimiento de esta condición significara una violación al Reglamento Interno en caso de empleados de la institución y una violación Legal de Propiedad Intelectual en caso de personas ajenas a la misma, dejando a discreción y derecho del Instituto Nacional de Estadística las acciones legales consiguientes.

	ELABORADO POR:	REVISADO POR:	APROBADO POR:
Nombre y Cargo:			
Cargo:			
Firma:			
Fecha:			

CONTROL DE EDICIONES

Versión:	Motivo de revisión	Modificaciones realizadas	Elaborado por:
Fecha:			

Versión:	Motivo de revisión	Modificaciones realizadas	Modificado por:
Fecha:			

Versión:	Motivo de revisión	Modificaciones realizadas	Modificado por:
Fecha:			

Instituto Nacional de Estadística

DOCUMENTO METODOLÓGICO

**SISTEMA CNPV
Escaneo Digital**

Versión 1.0

Autor: Mario Carranza

2013

CAPITULO I

1. INTRODUCCION

En este documento se describen la información técnica-metodológica del diseño del módulo Escaneo Digital de la Boleta Censal Sistema de Actualización Cartográfica.

Este módulo fue creado con el objetivo de crear trabajos de escaneo en el Flujo Kofax KTM una vez capturada la imagen Multitif en el módulo de Escaneo Documental.

2. METODOLOGIA

La metodología utilizada para el Módulo de Administración del Escaneo Digital es la de diseño y desarrollo rápido de un prototipo funcional para la interacción del usuario con el sistema, por la cual en base a la retroalimentación del mismo se refine el software. Gráficamente el modelo puede ser representado en la siguiente imagen:

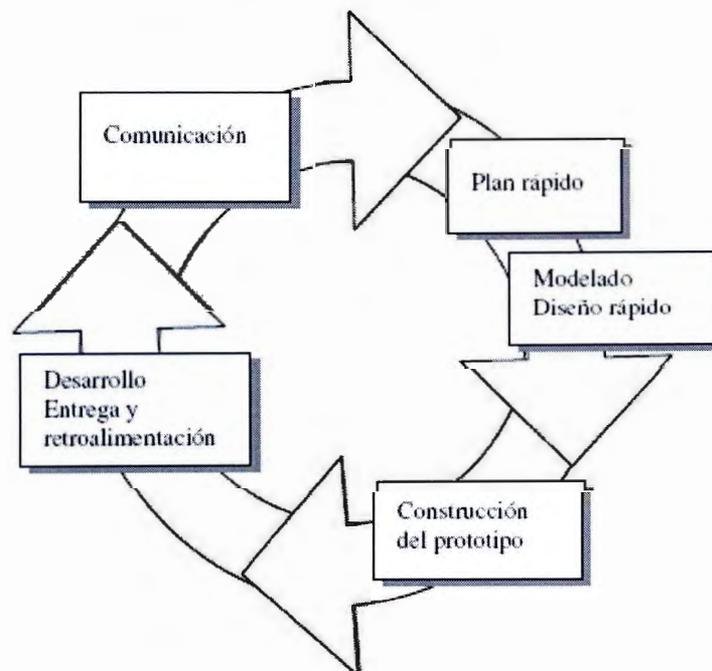


Imagen 1. Modelo de Construcción de Prototipo

3. HERRAMIENTAS

3.1 HERRAMIENTAS DE DISEÑO

UML (Lenguaje de Modelado Unificado), con su respectivos diagramas:

- Diagrama de Casos de Uso

3.2 HERRAMIENTAS DE DESARROLLO

Para el desarrollo Módulo de Escaneo Digital, se utiliza:

- Kofax Capture V. 10.0.
- Gestor de Base de Datos SQL Server 2012.

4. PROCESO Escaneo Digital

4.1 CASO DE USO GENERAL

En el módulo de Escaneo Digital, existen 6 procesos a seguir bien identificados (**Ver Imagen 2**):

- Creación de Trabajo
- Importación Imagen Multitif.
- Reconocimiento
- Revisión de Documentos
- Validación
- Exportación de Datos

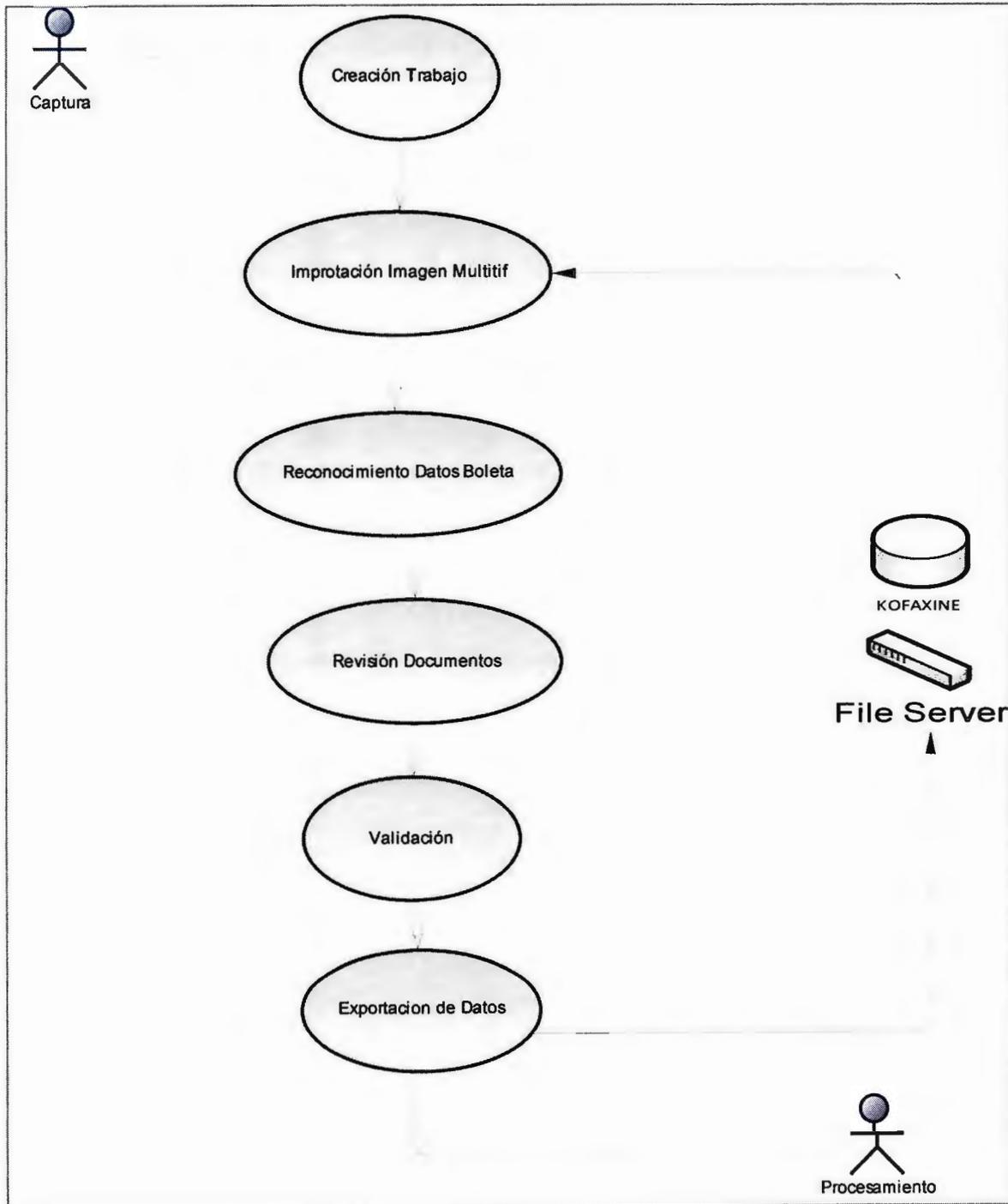


Imagen 2. Caso de uso general

4.2 CASOS DE USO EXPANDIDOS

4.2.1 CASO DE USO: Creación de Trabajo

Nombre de Caso de Uso: Creación de Trabajo Tipo: Primario	Actualizaciones: Original: Prioridad: Alta.
Actores: Usuario de Captura	
Descripción: <ul style="list-style-type: none"> • El usuario crea el trabajo mediante el módulo SCAN del Kofax, especificando: <ul style="list-style-type: none"> ○ Nombre ○ Carpeta ○ Turno ○ Carátula 	
Flujos alternativos: <ul style="list-style-type: none"> • El trabajo puede ser creado, en función a los requerimientos, de forma automática, en base a la información generada en el Módulo de Escaneo Documental. 	
Requisitos no funcionales: La información debe haber sido generada en el módulo de de Escaneo Documental.	
Precondiciones: Ninguna.	
Postcondiciones: <ul style="list-style-type: none"> • El trabajo es creado en la Base Interna de Kofax, dependiendo del servidor en el que se haya creado. 	
Diagrama de Actividades: No es necesario.	
Participantes: Usuarios de Captura de Datos, Procesamiento de Sistemas.	
Clases de análisis, responsabilidades, atributos: El usuario es responsable de los datos introducidos en el formulario.	
Prototipos de IU y reportes: No es necesario especificarlos en este caso.	

Tabla 1. Caso de uso: Creación de Trabajo

4.2.2 CASO DE USO: Importación Imagen Multitif

Nombre de Caso de Uso: Importación Imagen Multitif Tipo: Primario Actores: Usuario de Captura	Actualizaciones: Original: Prioridad: Alta.
Descripción: <ul style="list-style-type: none"> El usuario , una vez creado el trabajo de escaneo, importa el archivo Multitif creado en el Módulo Escaneo Documental, para su posterior reconocimiento. 	
Flujos alternativos: <ul style="list-style-type: none"> La importación puede realizarse, en función a los requerimientos, de forma automática, en base a la información generada en el Módulo de Escaneo Documental. 	
Requisitos no funcionales: El archivo debe haber sido creado en el módulo de de Escaneo Documental.	
Precondiciones: Ninguna.	
Postcondiciones: <ul style="list-style-type: none"> El archivo es cargado creado en la Base Interna de Kofax, dependiendo del servidor en el que se haya creado. 	
Diagrama de Actividades: No es necesario.	
Participantes: Usuarios de Captura de Datos, Procesamiento de Sistemas.	
Clases de análisis, responsabilidades, atributos: El usuario es responsable del archivo importado..	
Prototipos de IU y reportes: No es necesario especificarlos en este caso.	

Tabla 2. Caso de uso Importación Imagen Multitif

4.2.3 CASO DE USO: Reconocimiento Datos Boleta

Nombre de Caso de Uso: Reconocimiento Datos Boleta Tipo: Primario Actores: Usuario de Captura	Actualizaciones: Original: Prioridad: Alta.
Descripción:	

- Módulo Automático de Kofax que realiza el reconocimiento de ICR (Reconocimiento Inteligente de Caracteres, por su siglas en Inglés) y OMR (Reconocimiento Optico de Marcas, por sus siglas en inglés) de la Boleta Censal.

Flujos alternativos:

- Ninguno.

Requisitos no funcionales:

Ninguno.

Precondiciones:

Ninguna.

Postcondiciones:

- El trabajo pasa al Módulo de Revisión de Documentos o al Modulo de Control de Calidad.

Diagrama de Actividades:

No es necesario.

Participantes:

Usuarios de Captura de Datos, Procesamiento de Sistemas.

Clases de análisis, responsabilidades, atributos:

Prototipos de IU y reportes:

No es necesario especificarlos en este caso.

Tabla 3. Caso de uso: Reconocimiento Datos Boleta

4.2.4 CASO DE USO: Revisión de Documentos

<p>Nombre de Caso de Uso: Revisión de Documentos</p> <p>Tipo: Primario</p>	<p>Actualizaciones: Original: Prioridad: Alta.</p>
<p>Actores: Usuario de Captura</p>	
<p>Descripción:</p> <ul style="list-style-type: none"> • Módulo Asistido orientado a la correcta Calificación de Documentos de la Boleta: Carátula, Vivienda, Persona y Ultima Página. 	
<p>Flujos alternativos:</p> <ul style="list-style-type: none"> • Ninguno. 	
<p>Requisitos no funcionales: Ninguno.</p>	
<p>Precondiciones: Ninguna.</p>	
<p>Postcondiciones:</p> <ul style="list-style-type: none"> • El trabajo pasa al Módulo de Validación o al Modulo de Control de Calidad. 	
<p>Diagrama de Actividades:</p>	

No es necesario.

Participantes:

Usuarios de Captura de Datos, Procesamiento de Sistemas.

Clases de análisis, responsabilidades, atributos:

Prototipos de IU y reportes:

No es necesario especificarlos en este caso.

Tabla 4. Caso de uso: Revisión de Documentos

4.2.5 CASO DE USO: Validación

Nombre de Caso de Uso: Validación	Actualizaciones:
Tipo: Primario	Original:
Actores: Usuario de Captura	Prioridad: Alta.
Descripción:	
<ul style="list-style-type: none"> Módulo Asistido orientado a la validación de los campos manuscritos y/o burbujas de los Documentos de la Boleta: Carátula, Vivienda y Persona. 	
Flujos alternativos:	
<ul style="list-style-type: none"> Ninguno. 	
Requisitos no funcionales:	
Ninguno.	
Precondiciones:	
Ninguna.	
Postcondiciones:	
<ul style="list-style-type: none"> El trabajo pasa al Módulo de Exportación o al Modulo de Control de Calidad. 	
Diagrama de Actividades:	
No es necesario.	
Participantes:	
Usuarios de Captura de Datos, Procesamiento de Sistemas.	
Clases de análisis, responsabilidades, atributos:	
Prototipos de IU y reportes:	
No es necesario especificarlos en este caso.	

Tabla 5. Caso de uso: Validación

4.2.6 CASO DE USO: Exportación de Datos

Nombre de Caso de Uso: Exportación de Datos	Actualizaciones:
Tipo: Primario	Original:
	Prioridad: Alta.

Actores: Usuario de Captura	
Descripción:	<ul style="list-style-type: none"> • Módulo Automático de Kofax que realiza la exportación de Datos a las siguientes estructuras de Datos, en la Base de Datos KOFAXINE: <ul style="list-style-type: none"> ○ KFX_BOLETA_CARATULA_KTM ○ KFX_BOLETA_DOCUMENTOS_KTM ○ KFX_BOLETACENSALPERSONA_KTM ○ KFX_BOLETACENSALPERSONA_KTM2 ○ KFX_BOLETACENSALVIVIENDA_KTM ○ KFX_BOLETACENSALVIVIENDA_KTM2
Flujos alternativos:	<ul style="list-style-type: none"> • Ninguno.
Requisitos no funcionales:	Ninguno.
Precondiciones:	Ninguna.
Postcondiciones:	<ul style="list-style-type: none"> • El trabajo se elimina <i>una</i> vez que se han exportado los datos y las imágenes.
Diagrama de Actividades:	No es necesario.
Participantes:	Usuarios de Captura de Datos, Procesamiento de Sistemas.
Clases de análisis, responsabilidades, atributos:	
Prototipos de IU y reportes:	No es necesario especificarlos en este caso.

Tabla 3. Caso de uso: Reconocimiento Datos Boleta

5. DIAGRAMA ENTIDAD RELACION



Imagen 2. Diagrama ENTIDAD - RELACION



Instituto Nacional de Estadística

CENSO NACIONAL DE POBLACION Y VIVIENDA 2012

AREA DE: DIRECCIÓN DE INFORMÁTICA, CARTOGRAFÍA E
INFRAESTRUCTURA ESPACIAL
ETAPA: POST CENSO



CAPTURA DE DATOS

INTRODUCCIÓN CAPTURA DE DATOS

001 - V.1

2013



POST CENSO

FASE: CAPTURA DE DATOS

"INTRODUCCIÓN CAPTURA DE DATOS"

Control de Documentación

Código: INE-001-V1 Área Funcional: Dirección de Informática, Cartografía e Infraestructura Espacial

Prohibida su reproducción sin previa autorización. El incumplimiento de esta condición significara una violación al Reglamento Interno en caso de empleados de la institución y una violación Legal de Propiedad Intelectual en caso de personas ajenas a la misma, dejando a discreción y derecho del Instituto Nacional de Estadística las acciones legales consiguientes.

	ELABORADO POR:	REVISADO POR:	APROBADO POR:
Nombre y Cargo:			
Cargo:			
Firma:			
Fecha:			

CONTROL DE EDICIONES

Versión:	Motivo de revisión	Modificaciones realizadas	Elaborado por:
Fecha:			

Versión:	Motivo de revisión	Modificaciones realizadas	Modificado por:
Fecha:			

Versión:	Motivo de revisión	Modificaciones realizadas	Modificado por:
Fecha:			

CAPTURA DE DATOS

Personal de captura de datos

2013



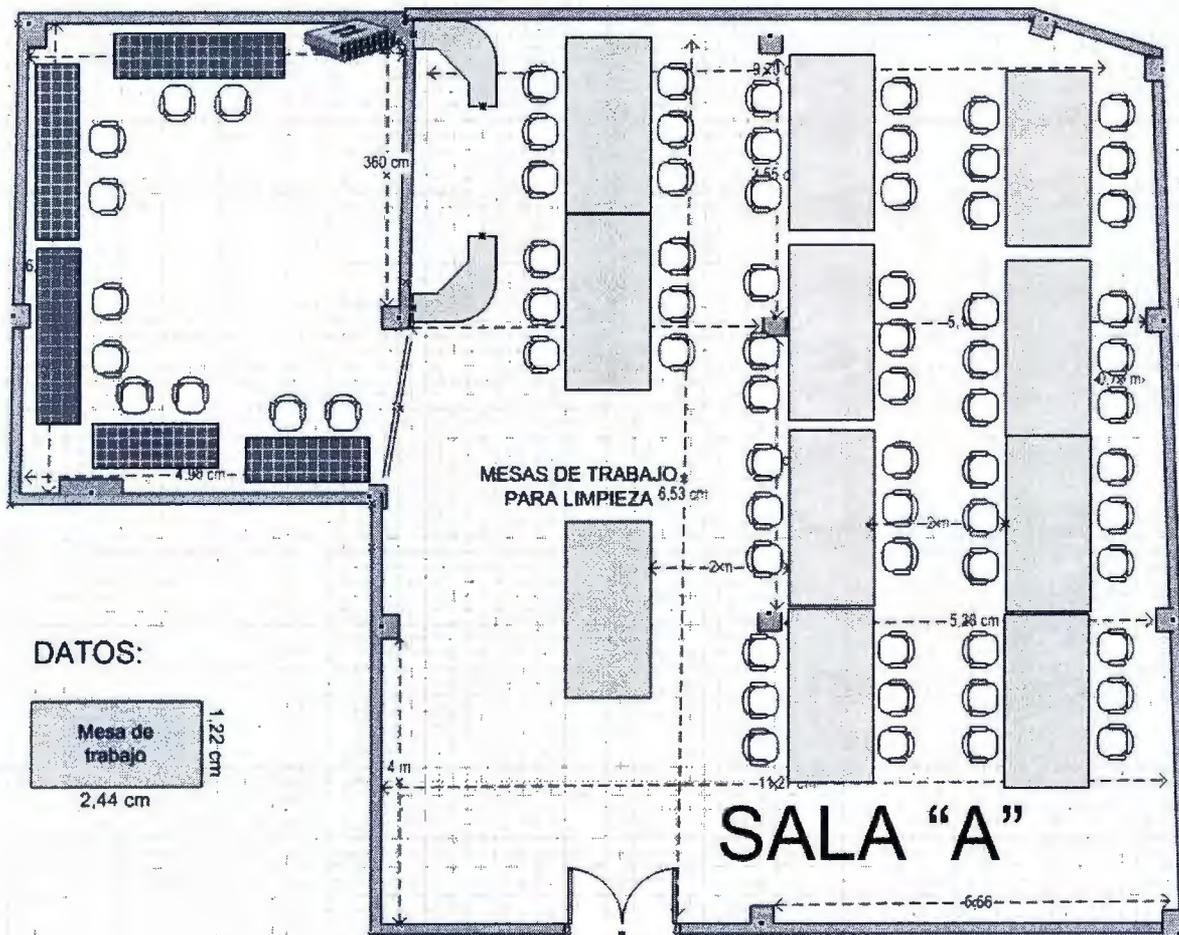
Este documento es un compendio de las tareas y procedimientos asignados al área de Captura de datos como parte de la Unidad de Procesamiento de datos. Así como la interacción con el resto de las áreas.

INSTITUTO NACIONAL DE ESTADÍSTICA
CARRASCO NRO. 1391
TEL. 2222333

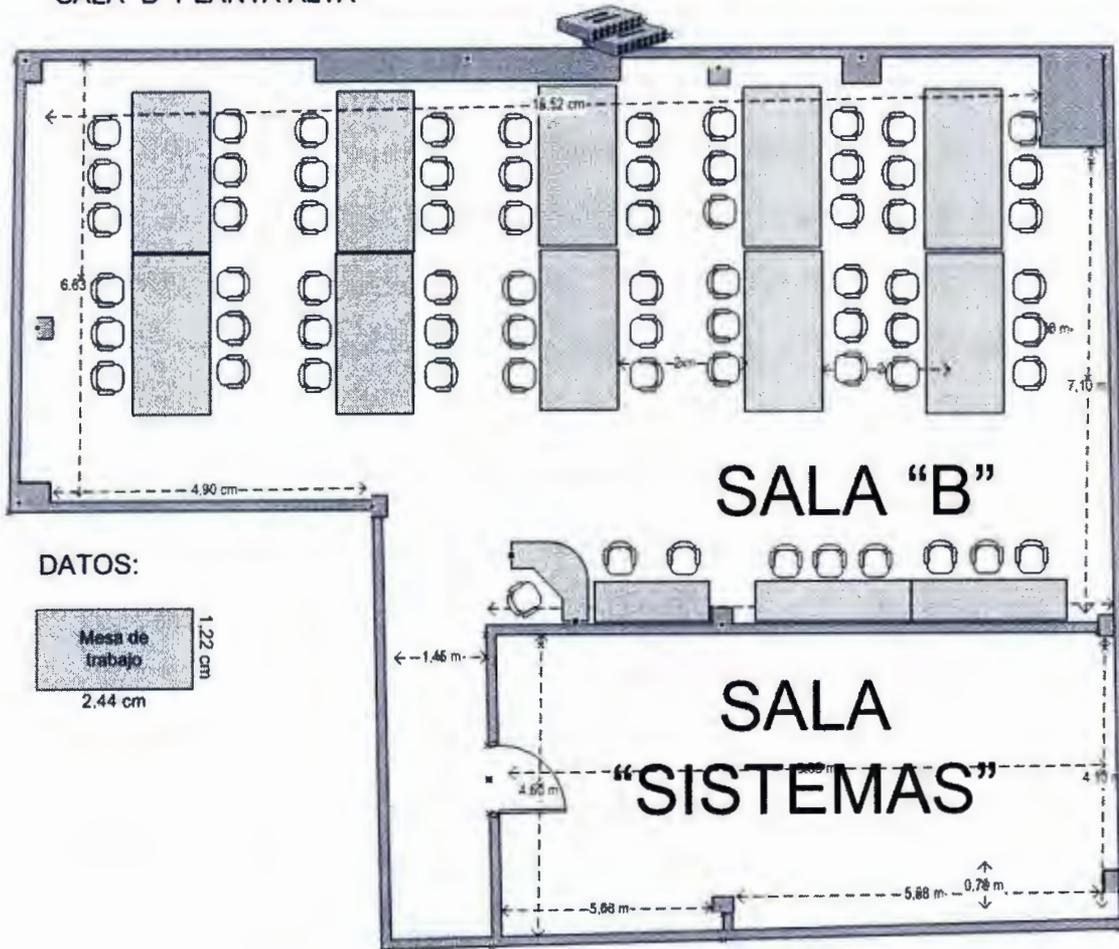
Inicialmente se preparó dos salas para la implementación de el escaneo y el personal de captura de datos.

En la Sala A en 5 mesas se instaló en enero la cantidad de 8 equipos de escaner para iniciar el trabajo del escaneado (se mantuvo 2 escáners para reserva), a principios de abril se habilitaron en la sala de C, a la que llamamos Sala de Escáner completando poco a poco a 25 escáners (con escáners prestados de otras instituciones y alquilados), para mayor información al área de Soporte Técnico.

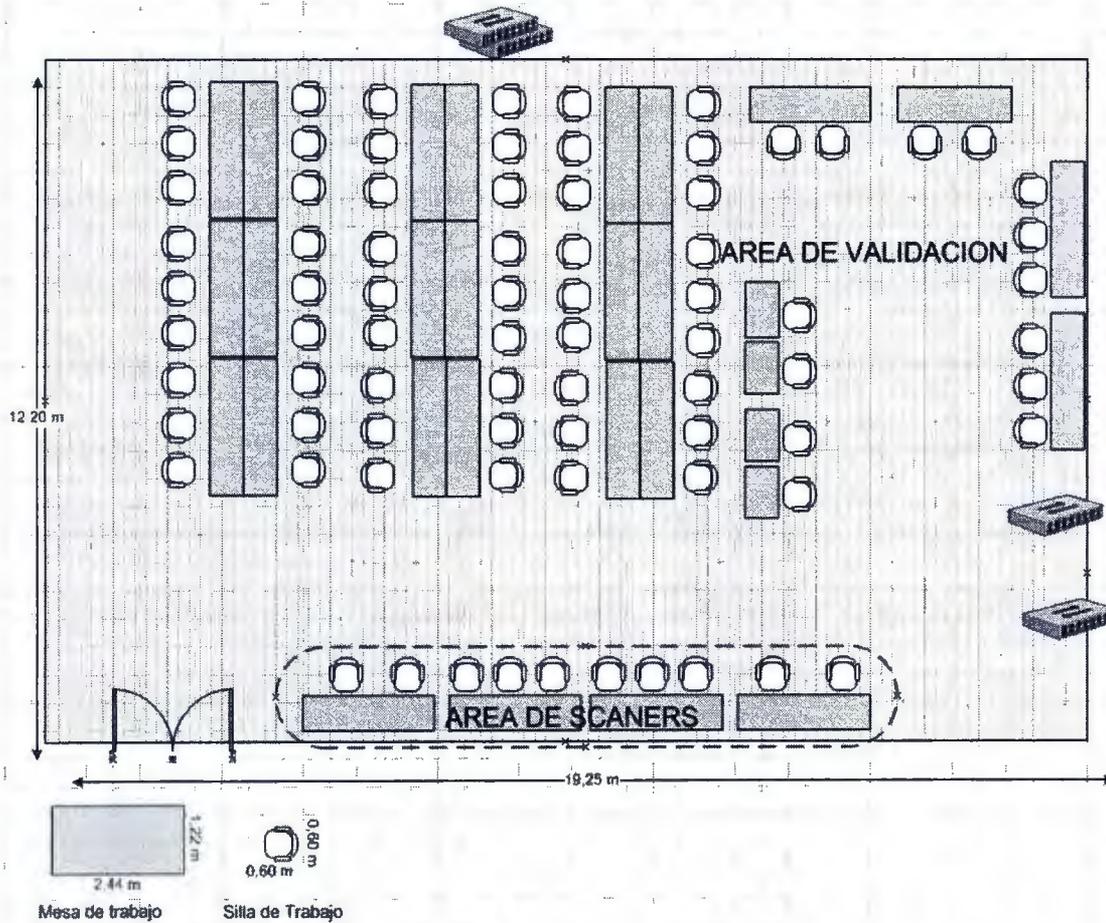
A continuación se presentan la configuración de las salas :



SALA "B" PLANTA ALTA



Para el mes de junio se adicionó más equipos en la Sala B y la sala A para que se pudiera incrementar al personal de transcripción-validación.



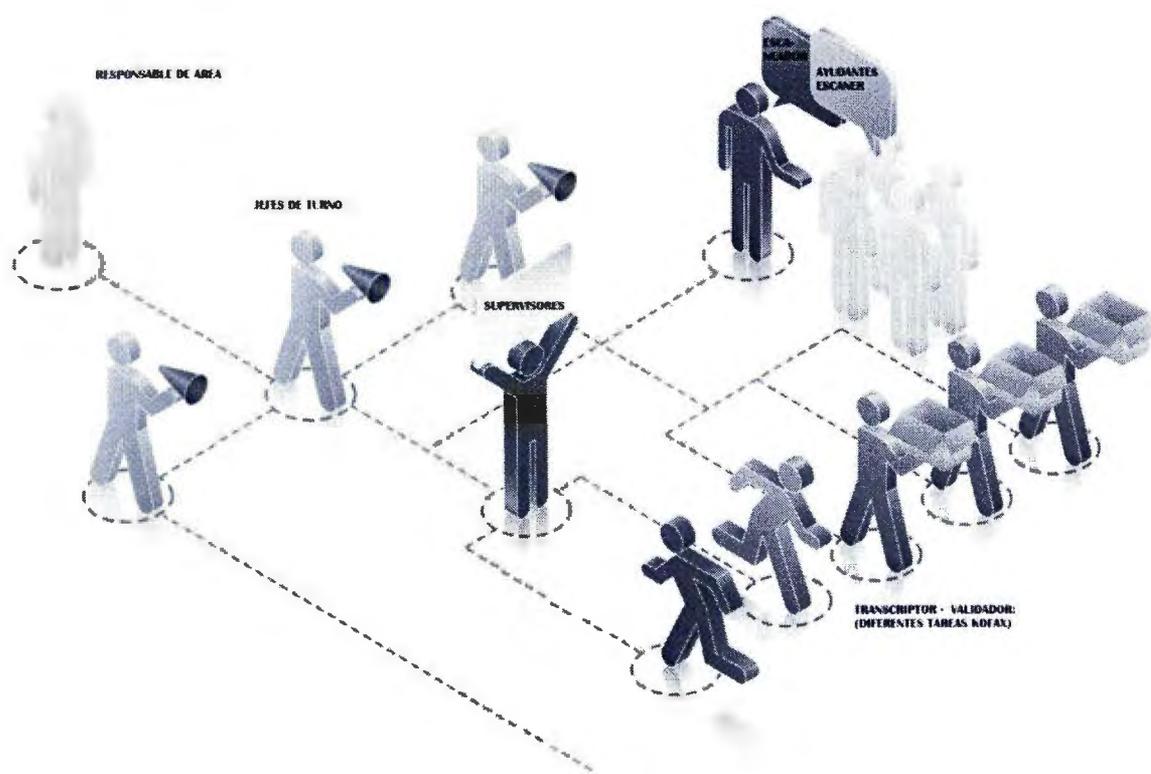
Se inicio con 237 transcriptoros - validadores, de ellos se quedaron hasta el final un poco más del 50%.

Se hizo una convocatoria en marzo para que inicien sus actividades en abril y el proceso fue un poco largo, ya que fue para completar los turnos diurnos y formar el turno nocturno.

Se realizaron adendas al personal que tuvo un rendimiento apreciable.

El turno nocturno fue el que más bajas tuvo y además por la necesidad de terminar los departamentos y por ello se realizó nuevos incrementos de personal.

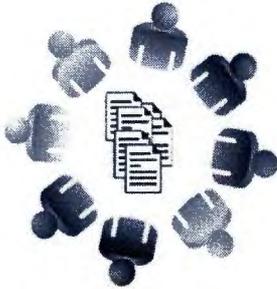
Lo nombrado anteriormente se refiere a infraestructura, equipos y personal, en lo que se refiere al proceso mismo, se presenta a continuación de forma escueta en el entorno que compete a Captura de Datos.



El trabajo inicia en la recepción de cajas censales, del área de Consistencia Básica, que pueden albergar una o más bolsas que contienen boletas identificadas con carátulas (ubicación geográfica), dentro de cada bolsa puede contener una o más carátulas que identifican una comunidad o manzano (unidades primarias).

En el área de Consistencia básica como su nombre lo indica realiza la revisión básica de las boletas censales de tal forma que se plasmen en el registro de la Base de Datos y la revisión de las boletas (estado físico) su ordenamiento y sea coherente con la carátula.

Una vez que se entrega al Área de Captura de datos se debe revisar el ordenamiento de las viviendas de forma secuencial, si contienen carpetas para que puedan pasar a la etapa de escaneo a esto le llamamos etapa de preescaneo (fue obligatorio esta revisión).



En la etapa de escaneo se registra en el sistema la identificación de la carátula que desde este momento se llamará "batch". Y posteriormente se introduce al sistema kofax tomando en cuenta las características de los perfiles para el escaneo que el área de Sistemas define. Estos perfiles van desde lo más sencillo como utilizar un determinado perfil para hojas transparentes de boletas, esto debido a la empresa que entregó las boletas censales impresas, hasta características más complejas de acuerdo al estado de la boleta (ejemplo mal guillotizadas).

Se creaba el batch por carátula, cabe recordar que en cada bolsa podían existir varias carátulas y esto quitaba tiempo al escaneador, posteriormente se vio la necesidad que exista el escaneo desatendido en el mes de abril, esto ayudó bastante pero luego se tuvo que revisar porque existía momentos en los que el sistema kofax no separaba automáticamente las carátulas y esto se debía realizar de forma manual.

Una vez que se escanean los batchs pasan a un proceso de exportación y posterior reconocimiento del batch en el kofax, que puede durar un tiempo corto o muy largo, dependiendo del servidor o la cantidad de baches paralelos que se escanean, mayor referencia al área de Sistemas, para luego pasar a la validación.

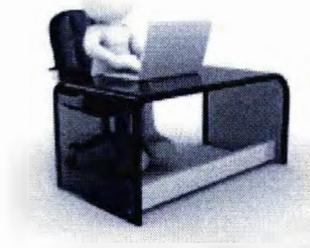


Dependiendo de las versiones KC o KTM los pasos para llegar a la validación han variado.

Se revisa el orden y clasificado por tipo de formulario (carátula, vivienda, persona)

Según lo propuesto por el área de Sis

Asignación de imágenes multitiff (KTM ,



En validación KC se valida todo y en el último mes se validó los campos determinados por Sistemas

La validación 1 en general las OMR y en validación 2 las ICR

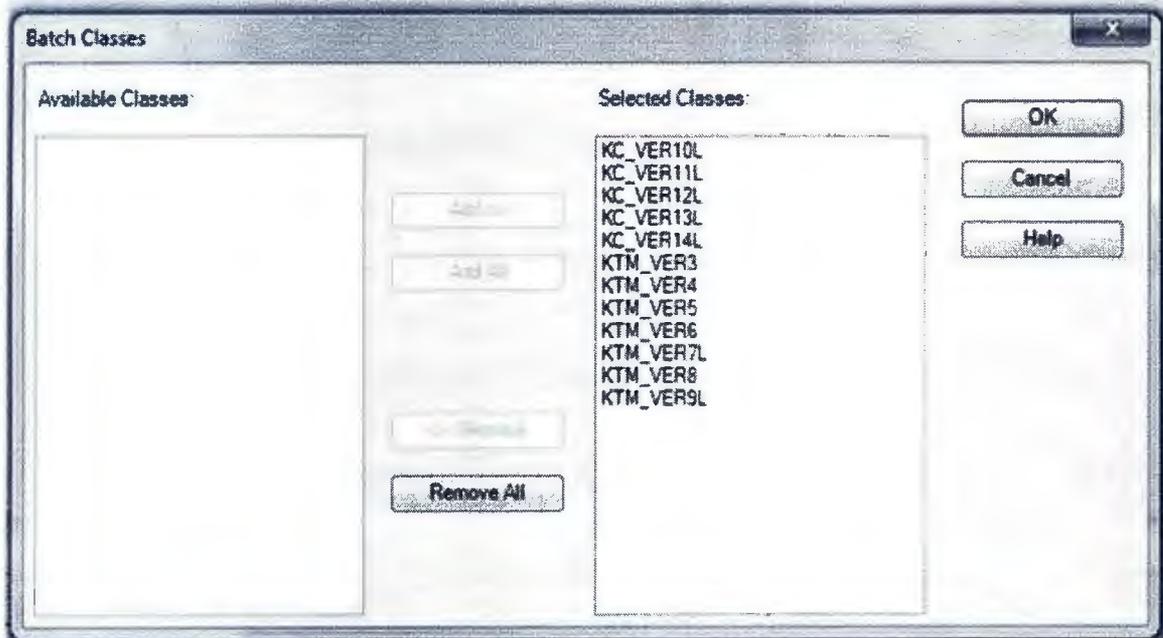
Document Review (KTM a partir de la versión 1) cuando se realizó la asignación automática en el mes de junio.



Utilizando el sistema se llama al archivo *.tiff según prioridad establecida por otra área (de forma manual), para introducir al kofax a través del módulo escáner, es un escaneado virtual (del físico) y se hace el nuevamente el reconocimiento del batch.

Validación KC solo una vez, Validación KTM que significa Validación 1 y validación 2





Quality control se dio en KC y KTM , este proceso es largo y se necesita de personal adiestrado y a la vez perspicaz.



Entra a quality control desde un cierre abrupto a medio validar, mala clasificación, etc hasta error de exportación

Proceso de retorno a validación o quality control, en cualquier momento puede volver a retornar a los pasos previos.

La etapa de generación del PDF y exportación corresponde exclusivamente al área de Sistemas. Pasos posteriores es de responsabilidad de Sistemas.

AVANCE DE LOS DEPARTAMENTOS

En enero se recibió la instrucción de apoyar a logística con 30 transcriutores - verificadores y consistencia básica con cuatro transcriutores - validadores, esto disminuyó la capacidad de avance.

Según los plazos establecidos el área de Captura de Datos ha ido respondiendo a las necesidades en función del personal y la dotación de equipos, cabe resaltar que los equipos no llegaron a tiempo, hasta el 20 de marzo se había concluido Oruro, Beni, Pando y La Paz se escaneó un poco más del 50% y 30% asignados para KTM.

En diferentes oportunidades el área de Captura de Datos tuvo que asumir tareas que no correspondían, como ser la revisión de boletas cuando el área de Consistencia Básica debía realizar dichas tareas y esto nuevamente disminuyó el avance esperado.

En lo que corresponde al Área de Sistemas (incluyendo siempre a la empresa DAMANET) el Área de Captura de datos ha estado sujeto a las diferentes versiones propuestas y las diferentes configuraciones de modo de trabajo que han establecido. Esto ha significado demora en tiempo porque estuvimos aplicando sus metodologías hasta que resulten funcionando de una mejor forma.

Se realizaron diferentes estrategias para cumplir lo encomendado y se han modificado dos veces los plazos de entrega de la información, añadiendo a esto se incrementó el personal de captura de datos y del mismo modo se adicionó el turno nocturno.

En resumen el avance se retrasó por lo siguiente:

- La tardanza de entrega de cajas debido a que la guillotina de la institución no era suficiente para abastecer al área de captura de datos.
- La empresa que se hizo cargo del guillotinado no realizó con el debido cuidado el guillotinado de las boletas y esto causó desfases en el reconocimiento por el software e hizo que se validara todos los campos.
- Adecuación de perfiles del escáner debido al tipo del papel y densidad del color de las boletas }
- Problemas de los escáners alquilados Cannon que no funcionó como se esperaba (se puede decir que un buen porcentaje de lo escaneado con estos equipos, se tuvo que reescanear).
- Revisión de boletas en orden y contrarrestando con la carátula del batch (esto debía realizarlo el área de Consistencia básica), limpieza de las boletas (por residuos del guillotinado) del total de las boletas antes del escaneo ✓
- Realización de tareas adicionales que no se ha previsto como ser Asignación de imágenes, document Review, quality control, validación 2 (solo se esperaba una sola validación) y repeticiones de pasos que el software no dejaba pasar y se tenía que consultar sólo con los personeros de DAMANET. En los últimos días del trabajo se tenían que esperar las asignaciones de cargas de batches y esto demoraba muchísimo según nos informaron. Simplificando el avance

de las etapas por las que pasa un batch en el kofax era incierto porque si bien avanzaba a una etapa posterior podía volver a una etapa previa.

- Por otra parte la exportación no se realizaba de una forma continua y parecería que afectaba la lentitud de la validación (el sistema se volvía lento) o porque se llenaba el servidor de imágenes y se colapsaba, mayor referencia al área de Sistemas.
- Se ha realizado el rescaneo de baches de los diferentes departamentos según orden del Área de Sistemas.
- La adecuación de las diferentes versiones del kofax para el personal

Además de ello se designaron "encargados departamentales" que su función era realizar un seguimiento de completitud de departamentos que significan una serie de tareas y además de verificar que se exporten a la Base de Datos.

Por parte del Área de Captura de datos se hizo el esfuerzo de trabajar de Lunes a Domingo sin espacios para lograr los objetivos iniciales. Es decir cumpliendo la validación final de todos los departamentos según el visto bueno del Área de Sistemas que al final dio la confirmación de la conclusión del trabajo.

Y se entregaron todas las cajas (bolsas - baches) al Área de logística.

Se adjunta información adicional.