

Captura y procesamiento de las Boletas Censales por medio del escaneo

La experiencia de Costa Rica

Jacqueline Castillo

Antecedentes

- En los anteriores Censos de Costa Rica la captura de datos se realizó mediante la digitación manual de los cuestionarios.
- En el Censo 2000 los resultados definitivos se brindaban 12 meses después de ejecutado el censo.
- Para el Censo 2011 con el objetivo de obtener resultados censales más oportunos y de mayor calidad, se realizó un proceso de investigación y análisis de diferentes alternativas para la captura de datos censales: Digitación numérica tradicional; Digitación alfanumérica con codificación a posteriori; Escaneo y PDA.
- Resultado: la opción más adecuada para el caso del Censo 2011 de Costa Rica es el uso del escaneo



¿Cómo funciona el Escaneo?

- Este proceso captura la imagen de los cuestionarios por lectoras (escaners) y luego esas imágenes pasan por una etapa de interpretación automática, con la transformación de ellas en marcas, números y caracteres.
- Cuando la interpretación no es posible por algún problema en el llenado de la boleta, el sistema las envía a una estación de verificación manual para resolver las dudas.
- Es una combinación de lectora de marcas (OMR, por *Optical Mark Recognition*), y lectora de caracteres, es decir, letras y números (OCR, por *Optical Character Recognition*).
- Esa combinación de técnicas (OMR y OCR) es conocida también como ICR (*Intelligent Character Recognition*), o como IMR (*Intelligent Mark Recognition*).



Acciones para disminuir los riesgos asociados al uso de escaner

- La falta de experiencia del INEC en la utilización del escáner en la captura de datos era un riesgo que se esperaba minimizar por medio de la contratación de una empresa especializada en el uso de esta tecnología en censos de población.
- Esta tecnología además debía ser probada en el Censo Piloto de junio de 2010. Si el producto era satisfactorio se continuaba con el proyecto.

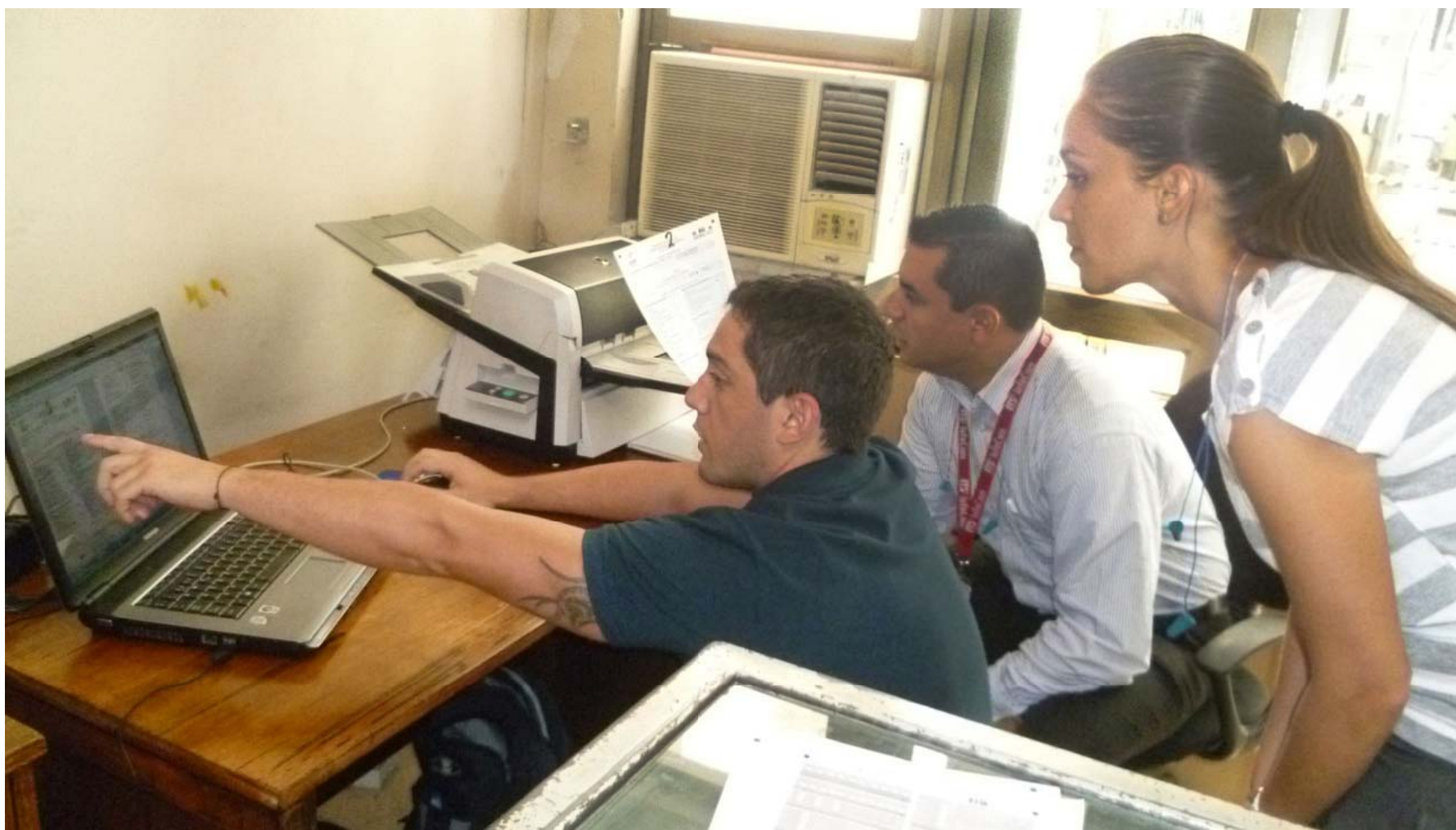


Contratación del Consorcio ITS-Active Software aportó una solución integral:

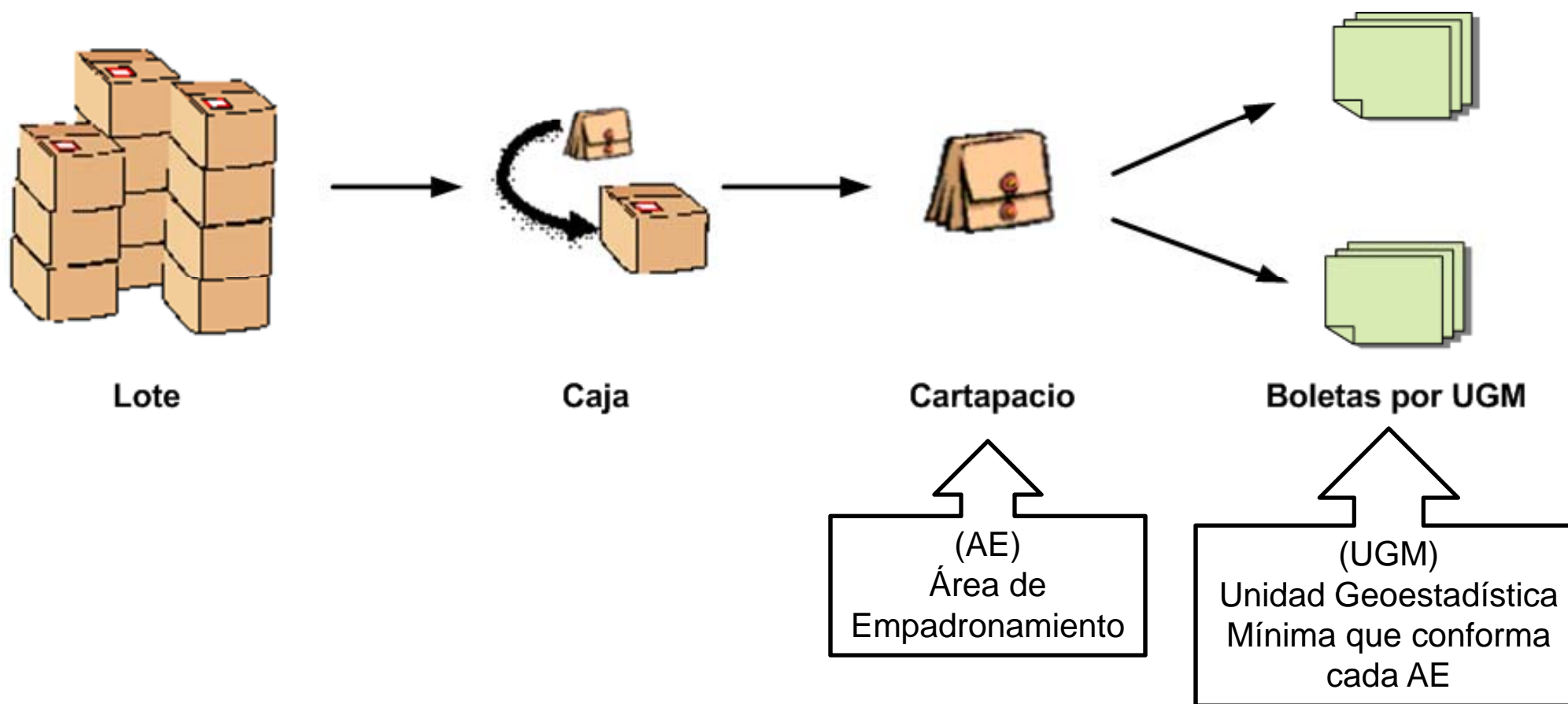
1. Captura de los datos del Censo Piloto 2010
2. Metodología de la Preparación y Procesamiento del Censo 2011
3. Diseño “fino” e Impresión de la Boleta Censal
4. La adquisición de los materiales asociados al llenado y manipulación de las boletas (lápices, borradores, tajador, cartapacio, cajas, bolsas de protección)
5. Armado o preparación de Cartapacios con las boletas
6. Distribución y Retiro del Material Censal en todo el país
7. Recepción y Almacenamiento del Material Censal
8. La captura de datos Censales y entrega de la Base de Datos final, incluyendo la codificación automática de preguntas abiertas.



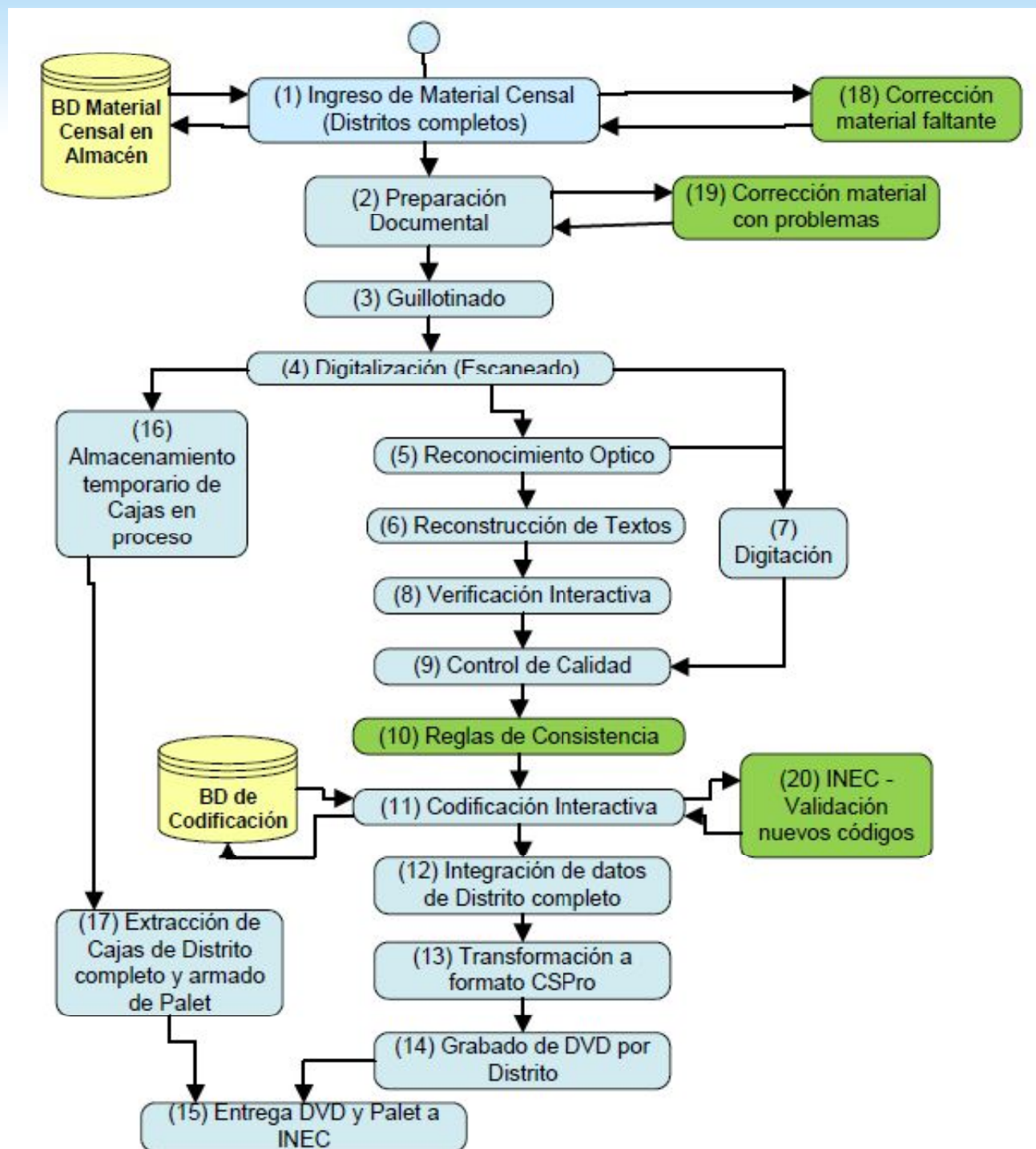
Proceso de captura y procesamiento de datos censales



Ordenamiento del material para realizar el proceso de captura



Proceso de escaneo y codificación



Consortio utilizó el Sistema Integral de Procesamiento Óptico (SIPO) para el seguimiento, documentación, conversión en datos y visualización de las boletas censales.

Sistema de Imágenes

Archivo Ver Herramientas

Criterio de búsqueda:

Formulario:

Etapas:

Consulta:

Buscar:

Lote	Carpeta	NroFormu...	ID_AE_C...	ID_UGM_1	ID_VIVIE...	ID_HOGA...	CONTINU...
0029	0012	0001	366070	185	1	0	
0029	0012	0002	366070	185	2	0	
0029	0012	0003	366070	185	3	1	
0029	0012	0004	366070	185	4	1	
0029	0012	0005	366070	185	5	1	
0029	0012	0006	366070	185	6	1	
0029	0012	0007	366070	185	7	1	
0029	0012	0008	366070	185	8	1	
0029	0012	0009	366070	185	9	1	
0029	0012	0010	366070	185	13	1	
0029	0012	0011	366070	185	14	1	
0029	0012	0012	366070	185	15	1	
0029	0012	0013	366070	185	16	0	
0029	0012	0014	366070	185	30	1	
0029	0012	0015	366070	185	17	1	
0029	0012	0016	366070	185	18	1	
0029	0012	0017	366070	185	20	1	
0029	0012	0018	366070	185	21	1	
0029	0012	0019	366070	185	22	1	
0029	0012	0020	366070	185	27	1	
0029	0012	0021	366070	185	31	1	
0029	0013	0001	366070	198	1	0	
0029	0013	0002	366070	198	2	1	
0029	0013	0003	366070	198	3	1	
0029	0013	0004	366070	198	4	1	
0029	0013	0005	366070	198	5	1	
0029	0013	0006	366070	198	6	1	
0029	0013	0007	366070	198	7	1	
0029	0013	0008	366070	198	8	1	
0029	0013	0009	366070	198	9	1	
0029	0013	0010	366070	198	11	1	
0029	0013	0011	366070	198	12	1	
0029	0013	0012	366070	198	13	0	
0029	0013	0013	366070	198	15	1	
0029	0013	0014	366070	198	16	1	
0029	0013	0015	366070	198	17	1	

SIPO

Sistema Integral de Procesamiento Óptico

Usuario - S.I.P.O.

Usuario:

Contraseña:



Características del proceso de escaneo para las imágenes

1. El sistema SIPO controla automáticamente que una Caja censal haya sido enteramente escaneada, ya que dispone de la información del contenido que dicha caja debería tener, o sea aquello que fue registrado en la recepción de cajas y sus UGM.
2. El operador de escáner trabaja por cajas enteras. El escáner no le permite comenzar con una nueva caja hasta que la caja en curso no esté terminada.
3. Respecto de la completitud de cada boleta, se verifica automáticamente que se hayan escaneado todas sus páginas, y en el orden requerido. El escáner no acepta faltantes ni hojas desordenadas.





BLOQUE I: LOCALIZACIÓN

Área de Empadronamiento

Unidad Geoestadística Mínima

Vivienda

Hogar

Si esta Boleta es CONTINUACIÓN
del mismo HOGAR, marque aquí:

Esta marca sólo aplica
a partir de la segunda
Boleta en hogares de
más de 6 personas.

197348 072 04 1

Dirección Residencial Abedules calle Laurel casa #25 E
Cipreses

Distrito Curridabat

INSTRUCCIONES DE LLENADO

ABCDEFGHIJKLMNÑOPQRSTUVWXYZ
0123456789

CORRECTO
Marque así

INCORRECTO
No marque así

BLOQUE II: CARACTERÍSTICAS DE LA VIVIENDA

1. Observe, indague y marque el tipo de vivienda:

INDIVIDUAL

Casa independiente 1
Casa independiente en condominio 2
Edificio de apartamentos 3
Edificio de apartamentos en condominio 4
Vivienda tradicional indígena (palenque o rancho) 5
Cuarto en cuartería 6
Tugurio 7
Otro (local, casa móvil, barco, camión) 8

COLECTIVA

Bache o barraca para trabajadores(as) 9
Albergue infantil 10
Hogar de personas adultas mayores 11
Cárcel 12
Otra (pensión, convento) 13

PERSONA SIN VIVIENDA

14

Pase a
BLOQUE III
Preg. 3

Pase a
BLOQUE IV

2. Indague y marque si la vivienda está ocupada o desocupada

OCUPADA

Con residentes habituales presentes 1
Con residentes habituales ausentes 2

DESOCUPADA

Para alquilar o vender 3
En construcción o reparación 4
Temporal para vacacionar 5
Temporal para trabajadores o estudiantes 6
Otra 7

Regresar
después

Fin de la
entrevista

3. ¿Esta vivienda...

... es propia totalmente pagada? 1
... es propia pagando a plazos? 2
... es alquilada? 3
... es prestada por motivo de trabajo? 4
... es prestada por otro motivo? (no paga) 5
... está en precario? 6
Otro 7

4. ¿Las paredes exteriores son principalmente de...

... block o ladrillo? 1
... zócalo? (cemento con madera o con fibrolit) 2
... madera? 3
... prefabricado o baldosas? 4
... Fibrolit, Ricalit? (lámina de fibrocemento) 5
... fibras naturales? (bambú, caña, chonta) 6
... material de desecho? 7
Otro (zinc, adobe) 8

5. ¿El techo es principalmente de...

... zinc? 1
... Fibrolit, Ricalit o asbesto? (fibrocemento) 2
... material natural? (palma, paja, suite u otros) 3
... material de desecho? 4
Otro (teja de arcilla, etc.) 5

6. ¿Tiene la vivienda cielo raso?

Sí 1 No 2

7. ¿El piso es principalmente de...

... cerámica, mosaico, terrazo? 1
... cemento? (ocre, lujado o no) 2
... madera? 3
... material natural? (chonta o rollizo) 4
Otro material 5
Piso de tierra 6

8. ¿El estado de...

	Malo?	Regular?	Bueno?
... las paredes exteriores es...	1	2	3
... el techo es	1	2	3
... el piso es	1	2	3

9. ¿Cuántos cuartos sólo para dormir tiene esta vivienda?

0 1 2 3 4 5 6 7 8 9 10
6 más

10. Incluyendo los cuartos ¿cuántos aposentos
en total tiene esta vivienda? (sin contar baños o cuarto de pilas)

1 2 3 4 5 6 7 8 9 10
6 más

Imágenes en formato TIFF (200 dpi)

1. Alta fidelidad para la digitalización de documentos
2. Amplio soporte de aplicaciones para la manipulación de imágenes
3. Calidad de imagen que brinda
4. Facilidad de compresión que se le puede dar sin perder calidad.



Reconocimiento automático de marcas, números y texto

Sistema de Imágenes

Imagen Imprimir Datos Cerrar

1

REPÚBLICA DE COSTA RICA
X CENSO NACIONAL DE POBLACIÓN Y VI DE VIVIENDA
censo 2011

BLOQUE I: LOCALIZACIÓN

Área de Emplacementamiento: 197455
Unidad Geoespacial Mínima: 094
Vivienda: 37
Hogar: 1

Si esta Boleta es CONTINUACIÓN del mismo HOGAR, marque aquí:

Esta marca sólo aplica a partir de la segunda Boleta en vigencia de más de 6 personas.

Dirección: San José, Curridabat, Cipreses, Condominio Colonia Del Roble
Distrito: Curridabat

INSTRUCCIONES DE LLENADO

ABCDEFGHIJKLMN OPQRSTUVWXYZ
0123456789

BLOQUE II: CARACTERÍSTICAS DE LA VIVIENDA

1. Observe, indague y marque el tipo de vivienda:

INDIVIDUAL

Casa independiente 1

Casa independiente en condominio 2

Edificio de apartamentos 3

Edificio de apartamentos en condominio 4

Vivienda tradicional indígena (palenque o rancho) 5

Cuarto en cuartería 6

Tugurio 7

Otro (local, casa móvil, rancho, camión) 8

COLECTIVA

Baño o barraca para habitadores(as) 9

Albergue infantil 10

Hogar de personas adultas mayores 11

Cárcel 12

Otra (penalón, convento) 13

PERSONA SIN VIVIENDA 14

2. Indague y marque si la vivienda está ocupada o desocupada:

OCUPADA

Con residentes habituales, presentes 1

Con residentes habituales, ausentes 2

DESOCUPADA

Para alquilar o vender 3

En construcción o reparación 4

Temporal para vacaciones 5

Temporal para trabajadores o estudiantes 6

Otra 7

3. ¿Esta vivienda...

... es propia totalmente pagada? 1

... es propia pagando a plazos? 2

... es alquilada? 3

... es prestada por motivo de trabajo? 4

... es prestada por otro motivo? (no pago) 5

... está en precario? 6

Otro 7

4. ¿Las paredes exteriores son principalmente de...

... block o ladrillo? 1

... zócalo? (cemento con madera o con fibra) 2

... madera? 3

... prefabricado o baldosas? 4

... Fibrolit, Ricalit? (láminas de fibrocemento) 5

... fibras naturales? (paja, caña, chonta) 6

... material de desecho? 7

Otro (zinc, alúbe) 8

5. ¿El techo es principalmente de...

... zinc? 1

... Fibrolit, Ricalit o asbesto? (fibrocemento) 2

... material natural? (ladrillo, paja, caña o roble) 3

... material de desecho? 4

Otro (paja de arroz, etc.) 5

6. ¿Tiene la vivienda cielo raso?

Si 1 No 2

7. ¿El piso es principalmente de...

... cerámica, mosaico, terrazo? 1

... cemento? (concreto o na) 2

... madera? 3

... material natural? (ladrillo o roble) 4

Otro material 5

Piso de tierra 6

8. ¿El estado de...

... las paredes exteriores es ... Mala? Regular? Bueno?

... el techo es 1 2 3

... el piso es 1 2 3

9. ¿Cuántos cuartos sólo para dormir tiene esta vivienda?

0 1 2 3 4 5 6 7 8 9 10 o más

10. Induciendo los cuartos ¿cuántos aposentos en total tiene esta vivienda? (se contar todas a cuarto de piso)

1 2 3 4 5 6 7 8 9 10 o más

Campo	Dato
LOTE	192
CARPETA	1
NÚMERO	37
PAGINA NRO	1
ID_AE_CONSECU_1	197455
ID_UGM_1	94
ID_VIVIENDA_1	37
ID_HOGAR_1	1
CONTINUACION_1	(sin datos)
V01_TIPO_VIVIENDA_1	2
V02_OCUPACION_1	4
V03_TENENCIA_1	(sin datos)
V04_MATERIAL_PAREDES_1	(sin datos)
V05_MATERIAL_TECNO_1	(sin datos)
V06_CIELO_RASO_1	(sin datos)
V07_MATERIAL_PISO_1	(sin datos)
V08A_ESTADO_PAREDES_1	(sin datos)
V08B_ESTADO_TECNO_1	(sin datos)
V08C_ESTADO_PISO_1	(sin datos)
V09_CANT_DORMITORIOS_1	(sin datos)
V10_CANT_APOSENTOS_1	(sin datos)

Base de datos

Sistema de Reconocimiento

Imagen

Verificación y codificación manual



Procedimientos de Control de Calidad de las imágenes y la base de datos

INEC verificó por muestreo:

1. Calidad de imágenes: comparación de imagen contra cuestionario papel
2. Calidad de la Base de datos: comparación de imagen contra Base de datos

Diariamente se recibían lotes procesados por lo que INEC realizó proceso paralelo y continuo de validación y aceptación de lotes (DVD con imágenes y datos)



Proceso de validación de datos realizado por INEC luego de recibido el lote



Resultados obtenidos

- Las actividades y productos entregados por el consorcio se recibieron satisfactoriamente.
- El procesamiento de datos censales se realizó en el tiempo esperado: 4 meses para la captura de cerca de 1.500.000 cuestionarios
- Conforme avanzaba el procesamiento se iban obteniendo resultados preliminares, facilitando los análisis de cobertura censal y consistencia de la información
- Se logró obtener una Base de Datos preliminar para el país al cabo de 4.5 meses.
- Se espera tener Base de Datos completa y definitiva 6 o 7 meses después del censo.



Recomendaciones

- Es indispensable identificar con detalle los procesos y actividades a desarrollar por la empresa antes de definir el cartel de contratación. Esto además aclara y define lo que deba hacer el INE y los recursos que necesita.
- El INE debe realizar supervisión directa del proceso, verificando el avance del trabajo y controlando la eficiencia y calidad de éste. Para lo cual debe asignarse el personal necesario.
- Lo anterior facilita la coordinación entre las partes y hace más oportuna la solución a los problemas que se presenten, buscando siempre la calidad en el producto a obtener.



- La experiencia de la empresa encargada de la captura de datos resulta determinante para obtener los resultados oportunos y de calidad
- La contratación de una solución integral evita muchos problemas administrativos y de logística asociados al censo (comunes en la administración pública)
- El INE debe realizar paralelamente las etapas de validación y consistencia de los datos (imputaciones) para garantizar que los resultados se obtengan oportunamente, lo que representa una alta exigencia para el personal.





censo 2011