



**UNIVERSIDAD AUTÓNOMA
DE AGUASCALIENTES**

**CENTRO DE CIENCIAS BÁSICAS
DEPARTAMENTO DE SISTEMAS DE INFORMACIÓN.**

“Propuesta de un Modelo de Calidad de Software para el Proceso de
Extracción, Transformación y Carga de Información Censal”

Trabajo de Investigación que para obtener el grado de

Maestro en Informática y Tecnologías

Computacionales

Presenta

Eduardo Esparza Juárez

Director de Tesis

MITC Jorge Eduardo Macías Luévano

Sinodales

Dra. Laura Garza González

MITC Cesar Eduardo Velázquez Amador

Cd. Universitaria, Mayo 2010

AGRADECIMIENTOS

*A Dios todo poderoso
por permitirme el don de la vida
de la que nunca se deja de aprender.*

*A mi esposa y mi amiga eterna Patricia Cuevas De Lira
por su apoyo, amor y comprensión en todo momento.*

*A mis hijos Patricia, Jesús Eduardo, Sebastián y Cristina Emilia
que sin saberlo me han permitido lograr una meta muy importante en mi vida profesional
y estoy seguro que un día cada uno de ellos superarán su desarrollo personal y profesional*

*A mis padres Soledad e Higinio
que desde que me enseñaron lo más elemental
me han acompañado en cada paso importante en mi vida*

*A mis hermanos Juan José, Armando, Octavio, Araceli e Higinio
pues siempre he encontrado su apoyo incondicional.*

*A mis maestros Jorge Eduardo, Laura y Cesar Eduardo
Director de tesis y revisores
por su apoyo y motivación durante mis estudios
y desarrollo de este trabajo*

*A mi maestra Cristinita Reyes Brand (descanse en paz),
nunca olvidaré sus primeras enseñanzas escolares.*



Centro de Ciencias Básicas

**L.I. EDUARDO ESPARZA JUÁREZ
PASANTE DE LA MAESTRÍA EN INFORMÁTICA
Y TECNOLOGÍAS COMPUTACIONALES
P R E S E N T E .**

Estimado (a) Alumno (a) Esparza:

Por medio de este conducto me permito comunicar a Usted que habiendo recibido los votos aprobatorios de los revisores de su trabajo de tesis y/o trabajo práctico titulado: **"PROPUESTA DE UN MODELO DE CALIDAD DE SOFTWARE PARA EL PROCESO DE EXTRACCIÓN, TRANSFORMACIÓN Y CARGA DE INFORMACIÓN CENSAL"**, hago de su conocimiento que puede imprimir dicho documento y continuar con los trámites para la presentación de su examen de grado.

Sin otro particular me permito saludarle muy afectuosamente.

A T E N T A M E N T E
Aguascalientes, Ags., 24 de junio de 2010
"LUMEN PROFERRE"
EL DECANO

DR. FRANCISCO JAVIER ALVAREZ RODRIGUEZ



c.c.p.- Archivo



Por este conducto autorizamos al tesista:

Eduardo Esparza Juárez

La impresión de su documento final de Tesis, ya que cumple con los requisitos de contenido y forma exigidos en la Universidad Autónoma de Aguascalientes.

Asesor



MITC. Jorge Eduardo Macías Luévano

Sinodales



Dra. Laura Garza González



MITC. César Eduardo Velázquez Amador

RESÚMEN.

Hablar de calidad en un producto implica necesariamente hablar de calidad en el proceso de manufactura de ese producto. La literatura reporta que las fallas durante el procesamiento de información a causa de la falta de un adecuado control de calidad han ocasionado fuertes pérdidas económicas, materiales, incluso organizacionales. En la industria del software se han hecho esfuerzos para atacar la problemática y generar proyectos de calidad. A raíz de estos esfuerzos es que han surgido propuestas de diversos modelos de procesos estándares que garantizan una mejora continua en el desarrollo del software. A pesar de los beneficios a nivel internacional por el uso de estos modelos, persiste la problemática pues estos modelos están fuertemente orientados a macro-organizaciones y proyectos de gran escala.

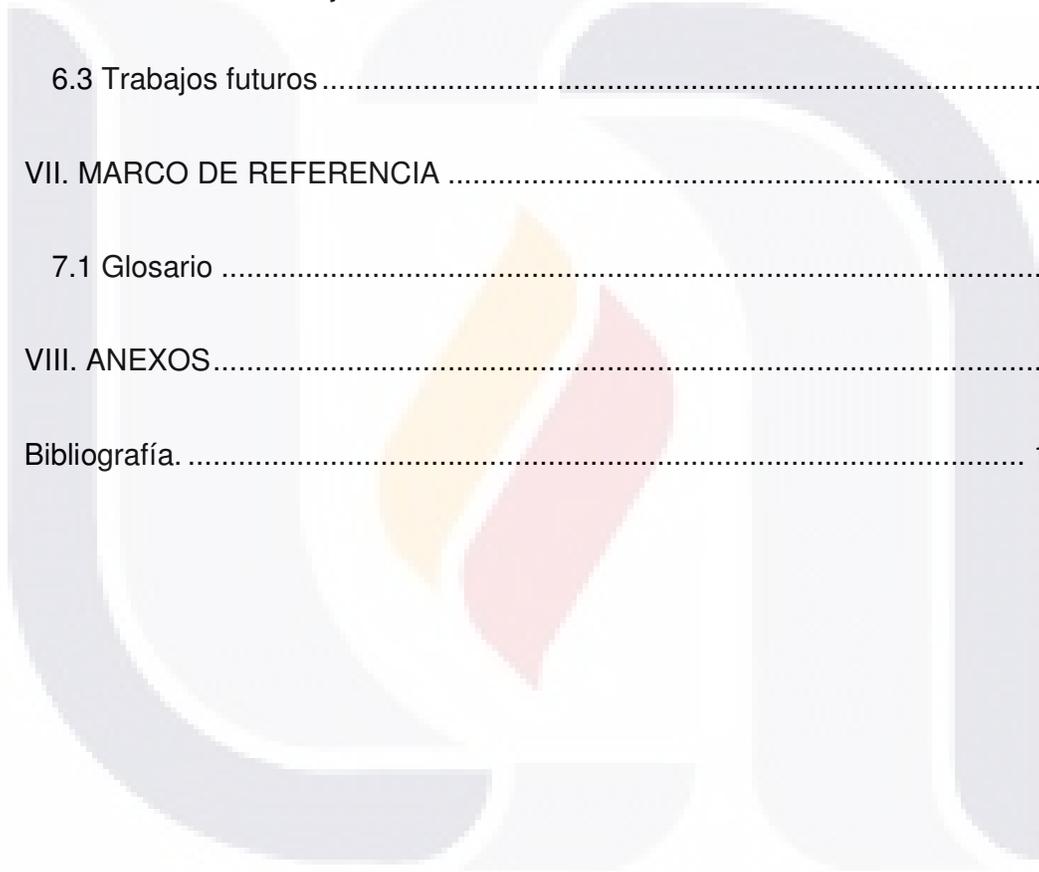
Este estudio pretende subsanar en parte esta problemática orientando los esfuerzos en particular al proceso de Extracción, Transformación y Carga específicamente sobre información censal debido a que necesariamente se refiere a grandes volúmenes. Para ello se toma como base la experiencia en dicho proceso por la institución que desarrolla los censos en México, el Instituto Nacional de Estadística y Geografía (INEGI) y la parte que abordan modelos usados por grandes compañías internacionales (CMMI, SPICE, MOPROSOFT)

De esta manera se obtiene la Propuesta de un Modelo de Calidad de Software para el proceso de Extracción, Transformación y Carga de Información Censal misma que contribuye al incremento de conocimiento científico en la línea de investigación general de la Ingeniería de Software y aporta las bases potencialmente útiles no solo al INEGI sino a los organismos o instituciones dedicadas a la generación de estadísticas basadas en información censal. Finalmente, en este estudio se presenta una propuesta para extender la investigación.

ÍNDICE

I.	Introducción.....	1
1.1	Descripción del entorno de desarrollo.....	1
1.2	Formulación del problema.....	4
II	JUSTIFICACIÓN.....	9
2.1	Justificación del problema.....	9
III.	Objetivos, Preguntas y Proposiciones.....	11
3.1	Definición de objetivos.....	11
3.2	Formulación de Preguntas de Investigación.....	11
3.3	Formulación de Proposiciones.....	12
3.4	Estudio de casos similares.....	12
3.5	Teorías y/o tecnologías utilizadas.....	16
IV.	MARCO TEÓRICO.....	17
4.1	Extracción, Transformación y Carga.....	17
4.2	Los procesos ETL de información censal sin modelo de calidad de software asociado.....	18
4.3	Modelos de calidad de software.....	72
V.	PROPUESTA DE MODELO DE CALIDAD DE SOFTWARE PARA EL PROCESO ETL DE INFORMACIÓN CENSAL.....	82

5.1 Definición.....	84
5.2 Bases prácticas.....	84
VI. CONCLUSIONES.....	92
6.1 Contribuciones, ventajas y desventajas de la propuesta.....	92
6.2 Evaluación de objetivos.....	93
6.3 Trabajos futuros.....	95
VII. MARCO DE REFERENCIA.....	96
7.1 Glosario.....	96
VIII. ANEXOS.....	97
Bibliografía.....	102



TESIS TESIS TESIS TESIS TESIS

I. Introducción.

1.1 Descripción del entorno de desarrollo

Organización actual.

El Instituto Nacional de Estadística, Geografía e Informática (INEGI) es un órgano desconcentrado de la Secretaría de Hacienda y Crédito Público, con autonomía técnica y administrativa. Para prestar el servicio público de información el INEGI cuenta con una Junta de Gobierno, un titular que representa a su Presidencia, una Contraloría Interna (CA) y Siete Direcciones Generales:

1. Dirección General de Coordinación del Sistemas Nacional de Información Estadística y Geográfica (DGCSNIEG)
2. Dirección General de Estadísticas Económicas (DGEE)
3. Dirección General de Estadísticas Sociodemográficas (DGES)
4. Dirección General de Geografía y Medio Ambiente (DGG)
5. Dirección General del Servicio Público de Información (DGSPI).
6. Dirección General de Vinculación Estratégica (DGVE)
7. Dirección General de Administración (DGA)

Para el ejercicio de sus funciones a nivel nacional, cuenta con una estructura descentralizada compuesta por diez Direcciones Regionales, que agrupan a 32 Coordinaciones Estatales y una Dirección General Adjunta de Apoyo Jurídico. (Intranet-INEGI. 2010)

Objetivos Institucionales

Los objetivos que permiten al Instituto dar cumplimiento a su plan de negocios son (Portal-INEGI. 2010):

- Generar e integrar información estadística y geográfica sobre el territorio, la población y la economía de México.
- Proporcionar a la sociedad el servicio público de información estadística y geográfica.
- Normar, coordinar y promover el desarrollo de los Sistemas Nacionales Estadístico y de Información Geográfica.
- Satisfacer las necesidades de información de los diversos sectores de la sociedad.

Misión, Visión y Políticas Institucionales

Para garantizar la mejora sustancial en la calidad y homogeneidad de la información, y además, hacer posible unir esfuerzos de las diferentes instancias y niveles de gobierno, en la integración de un sistema nacional, el INEGI se establece la siguiente misión, visión y políticas institucionales (Intranet-INEGI. 2010):

Misión. Generar, integrar y proporcionar información estadística y geográfica de interés nacional, así como normar, coordinar y promover el desarrollo de los Sistemas Nacionales Estadístico y de Información Geográfica, con objeto de satisfacer las necesidades de información de los diversos sectores de la sociedad.

Visión 2025. México pertenece al grupo de países que basan su desarrollo en el uso de la información y en el conocimiento organizado y diseminado electrónicamente al contar con un Sistema Nacional de Información Estadística y Geográfica sustentado en una Red Nacional de Información, que facilita la toma de decisiones de todos los sectores de la sociedad con base en información oportuna y confiable.

El INEGI es responsable de coordinar el Sistema Nacional de Información Estadística y Geográfica, así como la Red Nacional de Información.

Política de calidad. Todo producto o servicio que se genere en el INEGI debe tender a la plena satisfacción de las necesidades de información estadística y geográfica de la sociedad mexicana mediante el desarrollo de su personal y la mejora continua, privilegiando la integración de metodologías y tecnologías en sus procesos y proyectos.

Política Ambiental. El INEGI, coordinador y generador de información por medios electrónicos, ópticos e impresos, comprometido con el cumplimiento de la legislación ambiental vigente y otros requisitos que la organización suscriba, así como la prevención de la contaminación; establece para sus instalaciones los objetivos de promover el consumo responsable de materiales de oficina, energéticos, agua y el adecuado manejo de los residuos, procurando una mejora continua en el desempeño ambiental de sus tareas encomendadas.

Procesos clave

Como parte de su estrategia de negocios, el Instituto centra sus acciones en los siguientes procesos clave (Portal-INEGI. 2010):

- Establecer el marco normativo para la coordinación, desarrollo y funcionamiento de los Sistemas Nacionales Estadístico y de Información Geográfica.
- Promover el conocimiento y uso de la información estadística y geográfica, así como la integración de las instancias de participación previstas en la Ley de Información Estadística y Geográfica (LIEG) y vigilar su buen funcionamiento.
- Generar estadísticas de interés nacional con base en el levantamiento de censos, encuestas y el aprovechamiento de registros administrativos en materia económica, socio-demográfica, medio ambiente, ciencia y tecnología.
- Establecer las políticas, normas y técnicas para homogeneizar e integrar la información geográfica del país

- TESIS TESIS TESIS TESIS TESIS
- Establecer el Registro Nacional de Información Geográfica (RNIG) el cual contendrá, entre otra información, la división territorial del país y los límites internacionales, incluyendo la Zona Económica Exclusiva (ZEE).
 - Promover y coordinar tecnologías y metodologías que impulsen las competencias de los recursos humanos del Instituto.

1.2 Formulación del problema

Antecedentes de los censos de población y vivienda en México.

En la Ley del SNIEG se le atribuye al INEGI la facultad exclusiva de realizar los censos de población y vivienda, así como la responsabilidad de producir, difundir, promover y conservar la información estadística generada en todo el país, suministrar a la sociedad los resultados censales obtenidos de manera veraz y oportuna bajo los siguientes lineamientos básicos:

- La comparabilidad de la información en el tiempo y en el espacio.
- La adecuación conceptual, de acuerdo con las necesidades de la información que el desarrollo social y económico imponga.
- La comparabilidad internacional de la información que resulte de los censos mexicanos, atendiendo fundamentalmente a su periodicidad, con relación a la de otros países.

Desde el año 1900 los censos de población se realizan en México cada 10 años (excepto en 1920 cuando el país atravesaba por una revolución). En 1950 se introdujo el tema de vivienda como parte del censo y desde 1980 se identifican y caracterizan los hogares al interior de la vivienda. Posteriormente, en 1995 se realizó un evento muy similar al censo: El Conteo de Población y Vivienda, luego en 2005 el II Conteo de Población y Vivienda cuyo levantamiento de información se realizó durante octubre, gracias a la información obtenida por este evento fue posible la actualización de los principales indicadores sociodemográficos y contar con el máximo detalle de desagregación geográfica. Los eventos censales en el país se realizan sustentados legalmente en decretos publicados en el Diario Oficial de la Federación, a partir de esta publicación es que expertos al interior del INEGI

estudian y analizan los temas que deben ser abordados y la participación de las organizaciones gubernamentales, civiles y privadas (marco conceptual), así como el tamaño de la muestra en que se aplicará el o los cuestionarios censales, considerando para estas actividades las recomendaciones internacionales y los propósitos de comparabilidad estipulados en la propia Ley de Información Estadística y Geográfica.

Importancia del Censo de Población y Vivienda.

En la actualidad, el Censo de Población y Vivienda 2010 representa uno de los retos más importantes para el INEGI pues además de la importancia que de por sí está revestido se pretende implementar las tecnologías de información de vanguardia y personal profesional y de experiencia para lograr ahorros en tiempos de oportunidad de información, costos; la captación de información un porcentaje mayor de viviendas reduciendo el 4% de pendientes alcanzadas en operativos anteriores y la utilización de 7 medios diferentes de captación: Cuestionario Básico, Cuestionario Ampliado, Cuestionario de Autoenumeración, Cuestionario Electrónico (para captar información de viviendas colectivas), Cuestionario Electrónico (para captación de Autoenumeración), Cuestionario Electrónico (utilizado en el Centro de Atención Telefónica) y Cuestionario Internet (para captación de información de diplomáticos del Servicio Exterior Mexicano).

- Permite identificar los cambios que el país experimenta en sus aspectos demográficos, económicos y sociales, por lo que representa un soporte estadístico fundamental para apoyar la toma de decisiones de los sectores público, privado y social.
- Contribuye a respaldar la investigación que llevan a cabo académicos e instituciones educativas en el ámbito sociodemográfico.
- Fortalece el federalismo al proporcionar información comparable referida al ámbito estatal, municipal y local, posibilita la actualización de información que los organismos internacionales requieren con fines comparativos (ONU, OIT, CEPAL, CELADE, etcétera).

El o los cuestionarios censales son los instrumentos utilizados durante el levantamiento de información, es decir, contiene las preguntas que se aplican a la población a través de brigadas que visitan las viviendas para llenarlos. Una vez llenos los cuestionarios, se empaquetan y envían a las coordinaciones municipales (1306 a lo largo del país) aquí es donde se reciben y almacenan para luego proceder a su captura.

El procesamiento de la información considera 5 etapas fundamentales: Captura, Codificación, Validación, Actualización Cartográfica y Explotación, cada una de ellas entrega elementos que permiten el seguimiento y control de datos como salida e insumo entre las mismas.

Inmerso en este proceso de grandes volúmenes de información las operaciones de Extracción, Transformación y Carga son fundamentales y garantizar la calidad con que se realizan cobra niveles altos de importancia, por lo que se busca seguir un modelo de calidad de software que permita establecer una línea de acción y con ello generar información siempre confiable.

Situación problemática.

Mucho se ha escrito sobre los procesos de Extracción, Transformación y Carga (ETL por sus siglas en inglés) y los modelos de calidad de software en el desarrollo de sistemas de información en el área de ingeniería de software, sin embargo la intersección de estos temas tan abordados aplicado al proceso de generación de información censal puede resultar interesante y de utilidad en la aplicación de nuevos enfoques y modelos de software en los censos por venir.

Los procesos ETL que se han empleado para la información censal en los pasados años ha sido una práctica necesaria para consolidar lo que será la información de soporte a la generación de publicaciones y productos de consulta para la amplia gama de usuarios de ese tipo de información.

Además de ser una tarea necesaria también resulta ser una tarea en la que evidentemente se invierten recursos (humanos, materiales, económicos,

tiempo, etc.) y de la que se espera una calidad óptima para los procesos subsecuentes.

El presente trabajo pretende recurrir a los estudios sobre modelos de calidad de software en los procesos relacionados con la ETL y conjuntar con las experiencias tenidas en los conteos y censos de población desarrollados por el INEGI en México con el fin de proponer un nuevo modelo orientado a la información censal tomando como punto de partida las recomendaciones documentadas por los expertos en Ingeniería de Software.

Relevancia del caso de estudio.

La mayoría de los países contempla realizar un censo tradicional alrededor del año 2010. Varios de ellos experimentarán con nuevas tecnologías. Desean conocer los avances alcanzados por otros países de la región y fuera de ella que han adoptado metodologías alternativas.

El censo es parte del sistema integrado de la estadística socio-demográfica. En general, los INE producen información como insumos para apoyar la toma de decisiones la generación de nuevo conocimiento. [7]

La complejidad creciente, las nuevas condiciones sociales, y el potencial de las tecnologías, hacen imprescindible dar a las actividades gerenciales y estratégicas un nuevo contenido que junto con aprovechar la experiencia de censos anteriores, haga lo propio con los nuevos conocimientos y enfoques del área gerencial. [8]

Para realizar el censo se contratarán y capacitarán alrededor de 180 mil personas, de las cuales poco más de 106 mil se desempeñarán como entrevistadores; ellos tendrán un proceso amplio de capacitación para asegurar la calidad de la información en un primer instante, la captación.

De enero a mayo, se instalarán y equiparán más de 1 900 oficinas con acceso a Internet, entre otras actividades, los equipos instalados en estas oficinas tienen el objetivo de registrar de primera mano la información captada en 6 diferentes instrumentos de captación (cuestionario básico, cuestionario

ampliado, listado de inmuebles, cuestionario de posenumeración, etc.) y a través de distintos medios (cuestionario impreso en papel, cuestionario electrónico –vía internet o medios magnéticos-, etc.)

La experiencia y el conocimiento obtenido por eventos censales sustentan que los procesos de Extracción, Transformación y Carga (ETL) han resultado ser un cuello de botella para los sistemas que involucran estas etapas. Además el acceso a la información levantada en los operativos de campo y la información con la que se trabaja en las etapas siguientes cumple con cierto nivel de calidad que sigue un modelo de software llevado por la inercia de los expertos que han vivido y coordinado el tratamiento y la explotación de la información en los proyectos censales anteriores.

II JUSTIFICACIÓN.

2.1 Justificación del problema.

Beneficios del caso.

Con la propuesta del modelo de calidad de software en los procesos de Extracción, Transformación y Carga de información censal será posible la identificación oportuna de áreas de oportunidad y mejora continua implementando la tecnología, los procesos necesarios y las estrategias óptimas en las que se vierta la intervención de los tomadores de decisiones para retroalimentar y garantizar en conjunto la calidad de la información.

De esta manera el modelo de calidad de software para los procesos ETL constituye una base técnica que facilitará la claridad en la identificación de problemas con la información solventando las situaciones críticas que se derivan de los mismos.

Adicionalmente se recomienda asegurar la compatibilidad entre el modelo propuesto y los registros de seguimiento que como ya se mencionó cada proceso (captura, codificación, validación, actualización cartográfica y explotación) lleva a diferentes niveles de la información (segmento, municipio, entidad federativa).

Así mismo, la documentación y la sustentación de cada cambio que se realice a la luz del modelo de calidad propuesto permitirán considerarse en los proyectos censales siguientes.

Factibilidad del proyecto.

Para lograr la propuesta del modelo de calidad de software para el proceso de Extracción, Transformación y Carga existe literatura de referencia en la rama de la Ingeniería de Software para cada una de las etapas involucradas en la planeación, desarrollo e implementación de sistemas y/o aplicativos sobre las que se permitirá fundamentar el modelo de calidad propuesto.

Por lo tanto el modelo de calidad de software propuesto permitirá la mejora de los procesos ETL considerando el nivel de madurez que con respecto a la información censal ha alcanzado el INEGI y con ello mejorar la calidad para las etapas que da seguimiento la información hasta ser publicada.

En cuanto a la factibilidad económica y operacional, a través del procesamiento ETL sería posible determinar alguna situación anómala y se consideraría prioritaria cualquier mejora en los tiempos como parte del tratamiento a la información del Censo de Población y Vivienda 2010. Adicionalmente el proyecto censal está perfectamente alineado a las partidas presupuestales por lo que es factible destinar recursos técnicos y materiales para su desarrollo.

El proyecto deberá desarrollarse siguiendo el método de investigación para casos de estudio, que permitan tener una eficiente recolección de datos e insumos, un completo análisis de la información y una correcta interpretación de resultados, necesarios para proponer el modelo de calidad de software aplicado al proceso ETL de información censal.

Durante la etapa de pruebas del operativo de campo, el procesamiento de información se inicia con la captura de información en equipos laptop a través de aplicaciones stand-alone y conectados en red local hacia un servidor de BD con características propias en las oficinas de las Coordinaciones Municipales, de estas pequeñas BD se realizan envíos hacia una BD mayor en las oficinas centrales e integran en ella la información del país. Este paso de información de un oficina municipal a la central implica se procese el módulo ETL siguiendo el modelo de calidad de software propuesto.

III. Objetivos, Preguntas y Proposiciones

3.1 Definición de objetivos.

Objetivo General.

Proponer un modelo de calidad de software para el proceso de Extracción, Transformación y Carga (ETL por sus siglas en inglés) de información censal, a partir del análisis de diferentes modelos de calidad de software existentes y de la metodología que ha utilizado el INEGI en proyectos censales.

Particulares.

- Analizar la tendencia sobre los procesos ETL de diferentes modelos de calidad de software utilizados por pequeñas, medianas y grandes compañías.
- Analizar la metodología que ha utilizado el INEGI para procesar la ETL en los proyectos censales y determinar si la información generada es de calidad para las etapas posteriores del tratamiento.
- Proponer, como resultado del análisis, un modelo de calidad de software para el proceso de ETL de información censal.

3.2 Formulación de Preguntas de Investigación.

1. ¿Los modelos de calidad de software utilizados en pequeñas, medianas y grandes compañías orientan parte de sus procesos específicamente a la atención de los procesos ETL?
2. ¿La metodología utilizada por el INEGI en los eventos censales para el procesamiento ETL ha proporcionado información de calidad para las posteriores etapas del tratamiento?
3. ¿Es factible elaborar una propuesta de modelo de calidad para procesos ETL sobre información censal a partir de estudios de modelos de calidad existentes y/o la metodología utilizada por el INEGI?

3.3 Formulación de Proposiciones

1. Los modelos de calidad de software utilizados en pequeñas, medianas y grandes compañías están orientados a la mejora continua de sus procesos con el fin de obtener calidad en sus productos, pero la atención específica a los procesos ETL es mínima.
2. La información proporcionada por los procesos ETL en eventos censales realizados por el INEGI ha servido para el tratamiento de información en etapas posteriores pero no se ha evaluado su calidad.
3. Aunque en el estado del arte existen varios modelos de calidad de software con propósitos generales y particulares, a la fecha no han surgido propuestas de modelos de calidad de software orientados a los procesos ETL de información censal, pero es factible realizar una propuesta basada en los modelos existentes y la metodología utilizada por el INEGI en eventos censales.

3.4 Estudio de casos similares.

Como antecedente de la aplicación de los modelos de calidad de software orientados al proceso ETL se tienen los siguientes estudios.

a) Aseguramiento de la calidad de los datos en el proceso ETL del MNP [12].

El primer estudio consiste en el desarrollo de un sistema informático que permite el control de los flujos y la automatización de las tareas, asegura la calidad de los datos en todo el proceso de extracción, transformación y carga (ETL) realizado con los datos del Movimiento Natural de la Población (MNP). Su autor, el Centro Regional de Estadística de Murcia (CREM) indica que el control de los flujos, la gestión de permisos, la automatización de ciertos procesos, la existencia de tareas asistidas y la trazabilidad les permite evitar errores, mejorar la uniformidad en el tratamiento así como el control sobre los cambios producidos sobre sus bases de datos en cada momento. Además la obtención de informes con estadísticas del tratamiento de los datos les permite planificar sus procesos estadísticos de forma que se optimizan los recursos.

El trabajo presenta un software de interfaz gráfica apegada a los estándares de la intranet del CREM que facilita la gestión de administradores y usuarios del sistema asociados con un perfil que permite el acceso solo a funciones y datos determinados. Asimismo considera un flujo de control para dividir el trabajo considerando los estados de la información a través de las diferentes etapas por las que pasa.

De esta manera se exponen los factores que se controlan con el fin de aumentar la calidad de la información respecto a su exactitud, completitud y coherencia.

Las características de su proyecto que ayudan al aumento en la calidad de los datos destacan los siguientes aspectos:

- Flujo de control.
- Control total de los usuarios.
- Auditoria de control de cambios.
- Control de cambios de los valores posibles de los datos.
- Control de campos que pueden ser modificados.
- Generación de informes de rendimiento.
- Generación de informes sobre cada regla de validación.
- Permite el análisis de los cambios realizados por cada operador.
- Mayor autonomía del personal de estadística.

Muestra las alternativas tecnológicas que se tienen al alcance en el mercado para desarrollar el proyecto; y la justificación de la alternativa seleccionada.

Esquematiza el diagrama de flujo de información y de control de los procesos existentes en la actualidad.

En el país, al ser el INEGI la única institución encargada de realizar los censos de población y vivienda y ser la Ingeniería de Software un área de relativo reciente estudio, las indagatorias que se pudieran hacer sobre proyectos censales anteriores serían en base a la documentación realizada en los términos en que se haya generado en su momento. Sin embargo, es preciso mencionar que el presente trabajo busca dejar sentadas las bases para un

modelo de calidad sobre el proceso ETL y que además permita tener el precedente documentado y funja como el banco de conocimiento para los eventos censales por venir.

b) Nuevos métodos en la elaboración de estadísticas censales de población: EPV06 [16].

Este trabajo presentado por el organismo encargado de la producción de estadística oficial en la C. A. de Euskadi de la Unión Europea: EUSTAT, tiene como objetivo transmitir las experiencias en la realización de la Estadística de Población y Vivienda de 2006.

Inicialmente aclara que la generación de la información se realiza a partir de registros administrativos provenientes de varias oficinas y departamentos gubernamentales: Tesorería General de la Seguridad Social (TGSS), Mutualidad General de Funcionarios Civiles del Estado (MUFACE), Demandantes de empleo del Sistema Público de Empleo Estatal (SPEE/INEM), Departamento de Educación del Gobierno Vasco, Instituto Vasco de la Administración Pública (IVAP), Organismo Autónomo para la Euskaldunización y Alfabetización de Adultos (HABE), Diputaciones Forales, Departamento Vivienda y Asuntos Sociales, Departamento de Industria, Comercio y Turismo, ...) y de diversos productos estadísticos (Registro Estadístico de Población, Censo de Población y Viviendas de 2001, Población en Relación con la Actividad, Encuesta Demográfica y de Validación 2006, Censo del Mercado de Trabajo,...). Además del análisis de la información auxiliar procedente de la Encuesta Demográfica cuya muestra se compone de 30,000 viviendas con 25,000 de ellas ocupadas.

El autor señala que la estrategia de procesamiento inicia con la identificación, localización, petición y recepción de las fuentes de información a las cuales se les ha solicitado la información en archivos o bases de datos estructuradas de tal manera que facilite su integración a la base poblacional del Registro Estadístico de Población.

Los módulos informáticos desarrollados para el tratamiento de la información son:

- Aplicativo que realiza la carga de archivos, tratamiento de la información y la validación y corrección de las familias.
- Módulo de imputación del registro (MIR). Para el tratamiento estadístico de la información censal.
- Módulo de fusión (MDF). Para la fusión de registros.
- Módulo de codificación automática (CDR)

De estos módulos se pasa al detalle de cada uno de ellos, para el presente trabajo se considera relevante el primer módulo, en el que se dividen los tratamientos genéricos (que son los necesarios para la obtención de registros válidos mediante copia directa de variables) y específicos que se refieren a los que realizan conversiones necesarias de las variables seleccionadas.

Al interior de la descripción del primer módulo el autor muestra el diseño de las tablas de datos utilizadas así como las variables a las que se somete los tratamientos genéricos y el tratamiento que se da a las variables con procesamiento específico por cada tabla, es aquí donde se aplica el proceso de transformación para hacer compatible la información proveniente de las fuentes de datos con la base de procesamiento insumo de los siguientes módulos.

En el módulo de fusión se integran las fuentes de información con el entorno del Registro Estadístico de Población de Eustat, para este proceso se utilizan métodos automáticos y deterministas iniciando con las variables comunes posteriormente se implementaron métodos probabilistas. Se presentan informes como resultado del módulo.

Por último en el Módulo de Imputación del Registro describen las reglas que se consideran para imputar la información proveniente de la Encuesta Demográfica o de las fuentes administrativas, mostrando el número de registros imputados por cada regla.

3.5 Teorías y/o tecnologías utilizadas.

- ✓ Integración de Modelos de Madurez de Calidad (CMMI). Desarrollados por Paulk, Mark C. et al. de la Universidad de Carnegie-Mellon orientados a la mejora de procesos en diferentes niveles de madurez y hacia proyectos específicos.
- ✓ Moprosoft (Modelo de procesos para la industria de Software). Permite a las empresas mexicanas de desarrollo de software el acceso a las prácticas de Ingeniería de Software de clase mundial
- ✓ SPICE. Define procesos, técnicas y prácticas para la ingeniería de software [11].

IV. MARCO TEÓRICO

4.1 Extracción, Transformación y Carga

Definición

Es el proceso que organiza el flujo de los datos entre diferentes sistemas en una organización y aporta los métodos y herramientas necesarias para mover datos desde múltiples fuentes a un almacén de datos, reformatearlos, limpiarlos y cargarlos en otra base de datos, data mart o dataware house. [10]

Características.

- Identificación de información relevante por parte de la fuente de datos.
- Siguen reglas de transformación que se aplican a los datos antes de ser cargados al repositorio.
- Se define, describe, evalúa y documenta cada etapa del proceso.
- Existe un origen y un destino que pueden tener la misma o diferente plataforma.
- Las fuentes pueden ser extraídas completamente o aplicarse filtros.

Conceptos asociados

Workflow. Se refiere al orden de ejecución a partir del modelo conceptual al modelo lógico.

DataMart. Es una base de datos especializada en el almacenamiento de información de un área de negocio específica.

Almacén de datos. Es un gran conjunto de datamarts que proporciona al usuario una interfaz consolidada única para los datos y de esta manera facilitarles las consultas para la toma de decisiones.

ETL. Encargado de realizar las funciones de extracción de las fuentes de datos (transaccionales o externas), transformación (limpieza, consolidación principalmente) y la carga del Almacén de Datos, también hace el refresco del almacén (operación periódica que propaga los cambios de las fuentes externas al almacén de datos).

4.2 Los procesos ETL de información censal sin modelo de calidad de software asociado.

Captación de requerimientos

El requerimiento de los sistemas de software para el procesamiento de la información censal es emitido por la Dirección Técnica desde la Dirección General de Estadísticas Sociodemográficas hacia la Dirección de Informática en la Dirección General de Administración al interior del INEGI, de esta manera se integran las necesidades de procesamiento en cada una de las etapas por las que se da seguimiento y control a la información captada durante el censo. En cada etapa es necesario un proceso ETL para manipular la información como es requerido por cada una de ellas (Captura, Codificación, Validación, Actualización cartográfica y Explotación).

Identificación datos y fuentes de información.

Para la etapa de Captura la fuente de datos se encuentra directamente en los instrumentos de captación (cuestionarios) y la información se obtiene a través de diferentes medios: cuestionario impreso en papel, cuestionario electrónico (llenado vía web), CAT (Centro de Atención Telefónica), etc. Para el caso de los cuestionarios impresos en papel, éstos son capturados en las oficinas de las coordinaciones municipales instaladas en 1306 diferentes puntos a lo largo y ancho del país. En cada una de estas oficinas se instalaron aplicativos capturadores de información que registran la información en un pequeño motor de BD (SQL Lite).

La siguiente figura muestra el modelo de datos usado en las oficinas de las coordinaciones municipales donde se realiza la etapa de captura de la información censal:

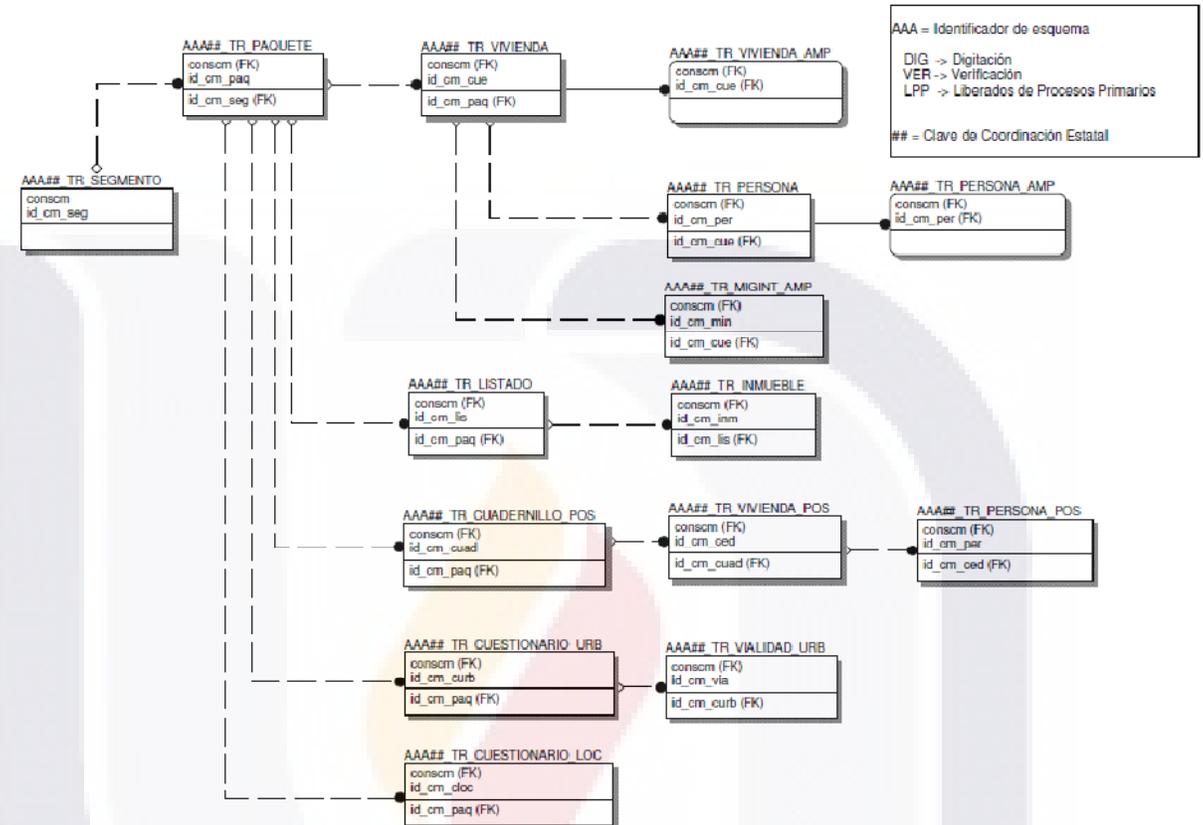


Fig.1 Modelo de datos en Coordinaciones Municipales

Localmente se transfieren los archivos a un servidor que a su vez realiza el envío a las oficinas centrales del INEGI en Aguascalientes mediante la aplicación denominada Administrador Local depositando dicha información en una estructura de directorios de un sitio FTP. Ver fig. 2.



Fig. 2 Transferencia de información capturada en Coordinaciones Estatales al sitio FTP en oficinas centrales.

Una vez que la información se encuentra en oficinas centrales, se realiza el primero y más importante de los procesos ETL el cual consiste en obtener los archivos del sitio FTP Extrayéndolos, Transformándolos y Cargándolos como datos a un servidor de base de datos (Oracle) dejando una estructura con información censal integrada de todas las oficinas de las coordinaciones municipales en un esquema nacional para base de datos.

La verificación que se realiza para saber si los datos que se tienen en la BD Oracle se hace a partir de controles propios del administrador de web que confronta cifras en cuanto a número de registros con los controles enviados por el administrador local. Además al interior de las tablas del modelo de datos se tiene registrado el número de paquetes, número de cuestionarios por paquete, número de personas al interior de cada cuestionario y número de cuestionarios por cada vivienda.

El modelo de datos para procesamiento de información en oficinas centrales cambia con respecto al utilizado en las coordinaciones municipales por ser un

manejo diferente el que se dará a la información (es por eso que fue necesaria la etapa de transformación de la información).

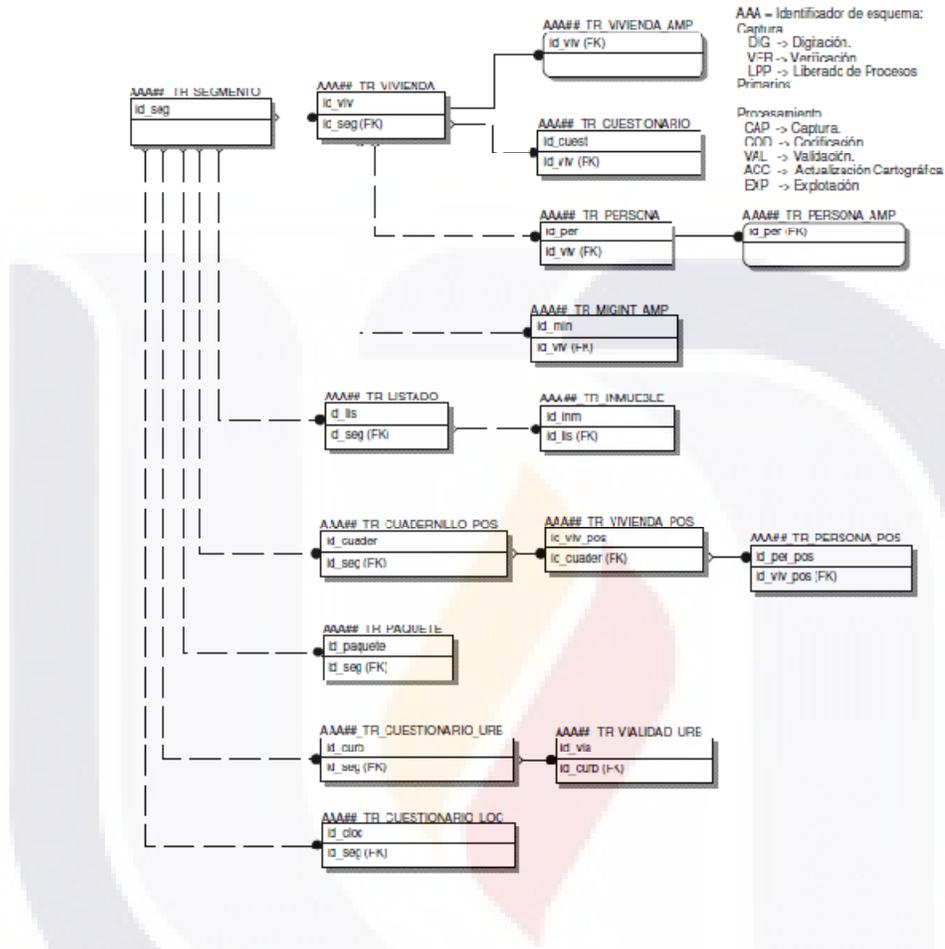


Fig. 3 Modelo de datos para procesamiento en oficinas centrales.

Descripción de los datos proporcionados por cada fuente de información.
 Las siguientes tablas muestran la información que se capta en cada una de las entidades identificadas en coordinaciones municipales.

CENSO DE POBLACIÓN Y VIVIENDA 2010
VARIABLES DE SALIDA DE CAPTURA POR TABLA

(captura tradicional)

CAP##_TR_SEGMENTS

Cons.	Descripción	Rango Válido
1	Consecutivo de CM	{1..9999}
2	Consecutivo de segmento en CM	{1..9999}
3	Identificación de Coordinación Estatal	{01..34}
4	Entidad Federativa	{01..32}
5	Municipio o Delegación	{001..570}
6	Clave de AGEB	{000..999;0..9 ó A,P}
7	Localidad	{0000..9999}
8	Manzana	{000..999}
9	Segmento	{A..I,N,P,Q,V}
10	Tipo de documento	{1,3..6}
11	Estado del segmento	{1..9}
12	Fecha de cierre de captura	{aaaammddhhmss}
13	Total de paquetes	{01..99,b}
14	Total de listados	{001..999,b}
15	Total de inmuebles	{001..999,b}
16	Total de viviendas	{001..999,b}
17	Total de viviendas ampliadas	{001..999,b}
18	Total de cuestionarios	{000..999,b}
19	Total de cuestionarios ampliados	{000..999,b}
20	Total de personas	{0000..9999,b}
21	Total de personas en cuestionario ampliado	{0000..9999,b}
22	Total de migrantes internacionales	{000..999,b}
23	Total de vías de acceso	{000..999,b}
24	Identificación de lote	{001..570;01..31;01..12,b}

CENSO DE POBLACIÓN Y VIVIENDA 2010
VARIABLES DE SALIDA DE CAPTURA POR TABLA

(captura tradicional)

CAP##_TR_PAQUETE

Cons.	Descripción	Rango Válido
1	Consecutivo de CM	{1..9999}
2	Consecutivo de paquete en CM	{1..9999}
3	Consecutivo de segmento en CM	{1..9999}
4	Entidad federativa	{01..32}
5	Municipio o delegación	{001..570}
6	AGEB	{000..999;0..9 ó A,P}
7	Localidad	{0001..9999}
8	Manzana	{000..999}
9	Segmento	{A..I,N,P,Q,V}
10	Folio de paquete	{101..999}
11	Tipo de documento	{1..6}
12	Consecutivo de paquete (operativo regular o especial)	{01..99,b}
13	Total de paquetes (operativo regular o especial)	{01..99,b}
14	Total de listados de inmuebles en el paquete (operativo regular o especial)	{00..99,b}
15	Total de documentos en el paquete (operativo regular o especial)	{000..999,b}
16	Área operativa (operativo regular o especial)	{000001..999999,b}
17	Área operativa (operativo regular o especial)	{0..9,b}
18	Consecutivo de paquete (operativo de verificación)	{01..99}
19	Total de paquetes (operativo de verificación)	{01..99}
20	Total de listados de inmuebles en el paquete (operativo de verificación)	{00..99,b}
21	Total de documentos en el paquete (operativo de verificación)	{000..999,b}
22	Área operativa (operativo de verificación)	{000001..999999,b}
23	Ubicación del paquete en almacén	{1-9999;A..Z;1..5;1..8}
24	Estado del paquete	{1..9}
25	Fecha de entrega al área de captura	{aaaammddhhmmss}
26	Fecha de entrega del área de captura al almacén	{aaaammddhhmmss}
27	Orden de captura	{1..9}

28	Clave del capturista	{00..99}
29	Fecha de captura	{aaaammddhhmmss}
30	Total de listados (captura)	{000..999,b}
31	Total de inmuebles (captura)	{000..999,b}
32	Total de cuestionarios (captura)	{000..999,b}
33	Total de cuestionarios ampliados (captura)	{000..999,b}
34	Total de personas (captura)	{0000..9999,b}
35	Total de personas en cuestionario ampliado (captura)	{0000..9999,b}
36	Total de migrantes internacionales (captura)	{000..999,b}
37	Total de vías de acceso para la manzana (captura)	{000..999,b}
38	Bandera de verificación	{1..4}
39	Total de viviendas	{01..99,b}
40	Total de cuestionarios	{000..999,b}
41	Campos con información	{00001..99999,b}
42	Promedio de campos con información por vivienda	{0001..9999,b}
43	Viviendas en muestra	{01..99,b}
44	Cuestionarios en muestra	{01..99,b}
45	Campos en muestra	{0000..9999,b}
46	Fracción de la muestra	{00.01...99.99,b}
47	Número aleatorio	{01..99,b}
48	Máximo de errores permitido	{01..99,b}
49	Errores de captura detectados	{00001..99999,b}
50	Consecutivos muestreados	{Alfanumérico,b}
51	Clave del verificador	{01..99}
52	Fecha de verificación	{aaaammddhhmmss,b}
53	Total de listados (verificación)	{000..999,b}
54	Total de inmuebles (verificación)	{000..999,b}
55	Total de cuestionarios (verificación)	{000..999,b}
56	Total de cuestionarios ampliados (verificación)	{000..999,b}
57	Total de personas (verificación)	{0000..9999,b}
58	Total de personas en cuestionario ampliado (verificación)	{0000..9999,b}
59	Total de migrantes internacionales (verificación)	{001..999,b}
60	Total de vías de acceso para la manzana (Verificación)	{001..999,b}

61	Fecha de procesamiento primario	{aaaammddhhmss,b}
62	Total de listados (LPP)	{000..999,b}
63	Total de inmuebles (LPP)	{000..999,b}
64	Total de cuestionarios (LPP)	{000..999,b}
65	Total de cuestionarios ampliados (LPP)	{000..999,b}
66	Total de personas (LPP)	{0000..9999,b}
67	Total de personas en cuestionario ampliado (LPP)	{0000..9999,b}
68	Total de migrantes internacionales (LPP)	{000..999,b}
69	Total de vías de acceso para la manzana (LPP)	{000..999,b}
70	Fecha de envío a Oficinas Centrales	{aaaammddhhmss,b}

CENSO DE POBLACIÓN Y VIVIENDA 2010
VARIABLES DE SALIDA DE CAPTURA POR TABLA
(captura tradicional)

CAP##_TR_LISTADO

Cons.	Descripción	Rango Válido
1	Consecutivo de CM	{1..9999}
2	Consecutivo de listado en CM	{1..9999}
3	Consecutivo de paquete en CM	{1..9999}
4	Entidad federativa	{01..32}
5	Municipio o delegación	{001..570}
6	AGEB	{001..999;0..9 ó A,P}
7	Localidad	{0001..9999}
8	Manzana	{000..999}
9	Segmento	{A..I,N,P,Q,V}
10	Apellido	{0..9}
11	Número de listado	{01..99}
12	Total de listados en la manzana, localidad o segmento	{01..99}
13	Tipo de cuestionario	{1..3,b}
14	Clave de entrevistador(a)	{0..9,b}
15	Clave de supervisor(a)	{0..9,b}
16	Clave de validador(a)	{0..9,b}
17	Clave de verificador(a)	{0..9,b}
18	Fecha de inicio de recorrido (día)	{00..99,b}
19	Fecha de inicio de recorrido (mes)	{00..99,b}
20	Fecha de término de recorrido (día)	{00..99,b}
21	Fecha de término de recorrido (mes)	{00..99,b}
22	Manzana o localidad	{1..3,b}
23	Tipo de documento	{1..4}
24	Bandera para liberar documentos sin información	{0,1}
25	Viviendas particulares con entrevista completa	{000..999,b}
26	Viviendas particulares con entrevista incompleta	{000..999,b}
27	Viviendas particulares con ausencia de residentes al momento de la vista	{000..999,b}
28	Viviendas particulares donde se negaron a dar información	{000..999,b}
29	Viviendas particulares deshabitadas	{000..999,b}
30	Viviendas particulares de uso temporal	{000..999,b}
31	Viviendas particulares donde no hay residentes habituales	{000..999,b}
32	Total de viviendas particulares	{000..999,b}

33	Viviendas con entrevista completa	{000..999,b}
34	Viviendas con entrevista incompleta	{000..999,b}
35	Viviendas con ausencias de residentes habituales al momento de la vista	{000..999,b}
36	Viviendas donde se negaron a dar información	{000..999,b}
37	Viviendas deshabitadas	{000..999,b}
38	Viviendas de uso temporal	{000..999,b}
39	Viviendas donde no hay residentes habituales	{000..999,b}
40	Total de viviendas	{000..999,b}
41	Viviendas colectivas con entrevista completa	{000..999,b}
42	Viviendas colectivas con entrevista incompleta	{000..999,b}
43	Viviendas colectivas con ausencias de residentes habituales al momento de la vista	{000..999,b}
44	Viviendas colectivas donde se negaron a dar información	{000..999,b}
45	Total de viviendas colectivas	{000..999,b}

CENSO DE POBLACIÓN Y VIVIENDA 2010

VARIABLES DE SALIDA DE CAPTURA POR TABLA

(captura tradicional)

CAP##_TR_INMUEBLE

Cons.	Descripción	Rango Válido
1	Consecutivo de CM	{1..9999}
2	Identificación de inmueble en CM	{1..99999}
3	Consecutivo de listado en CM	{1..9999}
4	Número de renglón (automático)	{01..20}
5	Código de confronta	{1..3,b}
6	Cotejador	{001..999,b}
7	Consecutivo del inmueble	{001..999,b}
8	Dirección del inmueble (calle, avenida, calzada, privada, carretera, andador)	{Alfanumérico,b}
9	Número visible (exterior)	{Alfanumérico,b}
10	Número visible (interior)	{Alfanumérico,b}
11	Total de viviendas en el predio	{000..999,b}
12	Total de locales en el predio	{000..999,b}
13	Verificación de residentes	{1,3,b}
14	Tipo de inmueble	{1..8,b}
15	Resultado de la visita (primera)	{1..7,b}
16	Resultado de la visita (segunda)	{1..7,b}
17	Resultado de la visita (tercera)	{1..7,b}
18	Resultado de la visita (supervisor)	{1..7,b}
19	Resultado de la visita (final)	{1..7,b}
20	Registro para localizar a los residentes	{Alfanumérico,b}
21	Código de verificación	{1..7,b}

CENSO DE POBLACIÓN Y VIVIENDA 2010

VARIABLES DE SALIDA DE CAPTURA POR TABLA

(captura tradicional)

CAP##_TR_VIVIENDA

Cons.	Descripción	Rango Válido
1	Consecutivo de CM	{1..9999}
2	Consecutivo de cuestionario en CM	{1..99999}
3	Consecutivo de paquete en CM	{1..9999}
4	Identificador único de vivienda	{1..99999999}
5	Entidad federativa	{01..32}
6	Municipio o delegación	{001..570}
7	AGEB	{001..999;0..9 ó A,P}
8	Localidad	{0001..9999}
9	Manzana	{000..999}
10	Segmento	{A..I,N,P,Q,V}
11	Apellido	{0..9}
12	Consecutivo de la vivienda	{000..999,b}
13	Número de cuestionario en la vivienda	{0001..9999}
14	Total de cuestionarios en la vivienda	{0001..9999}
15	Fecha de entrevista (día)	{00..99,b}
16	Fecha de entrevista (mes)	{00..99,b}
17	Clave de entrevistador(a)	{0..9,b}
18	Clave de supervisor(a)	{0..9,b}
19	Clave de validador(a)	{0..9,b}
20	Clave de verificador(a)	{0..9,b}
21	Clase de vivienda particular	{1..7,b}
22	Clase de vivienda colectiva	{Alfanumérico,b}
23	Folio de paquete	{101..999}
24	Consecutivo del cuestionario en el paquete	{001..999}
25	Total de ocupantes	{01..99}
26	Tipo de documento	{1..3}
27	Medio de captación	{1..7}
28	Bandera para liberar documentos sin información	{0,1}

29	Material predominante en pisos	{1..3,b}
30	Número de cuartos dormitorios	{00..99,b}
31	Total de cuartos	{00..99,b}
32	Electricidad	{1..3,b}
33	Disponibilidad de agua	{1..6,b}
34	Excusado	{1..3,b}
35	Admisión de agua	{5..7,b}
36	Drenaje	{1..5,b}
37	Bienes y TIC (Radio)	{1,2,b}
38	Bienes y TIC (Televisor)	{3,4,b}
39	Bienes y TIC (Refrigerador)	{1,2,b}
40	Bienes y TIC (Lavadora)	{3,4,b}
41	Bienes y TIC (Automóvil o camioneta)	{1,2,b}
42	Bienes y TIC (Computadora)	{3,4,b}
43	Bienes y TIC (Línea telefónica fija)	{1,2,b}
44	Bienes y TIC (Teléfono celular)	{3,4,b}
45	Bienes y TIC (Internet)	{1,2,b}
46	Número del informante	{00..99,b}
47	Número de personas en la vivienda	{00000..99999,b}
48	Verificación de la lista	{1,3,b}

CENSO DE POBLACIÓN Y VIVIENDA 2010

VARIABLES DE SALIDA DE CAPTURA POR TABLA

(captura tradicional)

CAP##_TR_VIVIENDA_AMP

Cons.	Descripción	Rango Válido
1	Consecutivo de CM	{1..9999}
2	Consecutivo de cuestionario en CM	{1..99999}
3	Identificador único de vivienda	{1..99999999}
4	Paredes	{1..8,b}
5	Techos	{1..9,b}
6	Cocina	{1,3,b}
7	Dotación de agua	{1..5,b}
8	Uso exclusivo	{1,3,b}
9	Combustible	{1..6,b}
10	Eliminación de basura	{1..8,b}
11	Tenencia	{1..3,b}
12	Adquisición	{1..4,b}
13	Equipamiento (estufa de gas)	{1,2,b}
14	Equipamiento (estufa de leña o carbón con chimenea)	{3,4,b}
15	Equipamiento (tinaco)	{1,2,b}
16	Equipamiento (calentador de agua)	{3,4,b}
17	Equipamiento (cisterna o aljibe)	{1,2,b}
18	Equipamiento (regadera)	{3,4,b}
19	Equipamiento (medidor de luz)	{1,2,b}
20	Condición de migración internacional	{1,3,b}
21	Número de personas	{00..99,b}
22	Acceso a la alimentación (¿alguna persona ... sólo comió una vez al día?)	{1,2,b}
23	Acceso a la alimentación (¿alguna persona ... dejó de comer todo un día?)	{3,4,b}
24	Acceso a la alimentación (¿alguna vez se quedaron sin comida?)	{1,2,b}

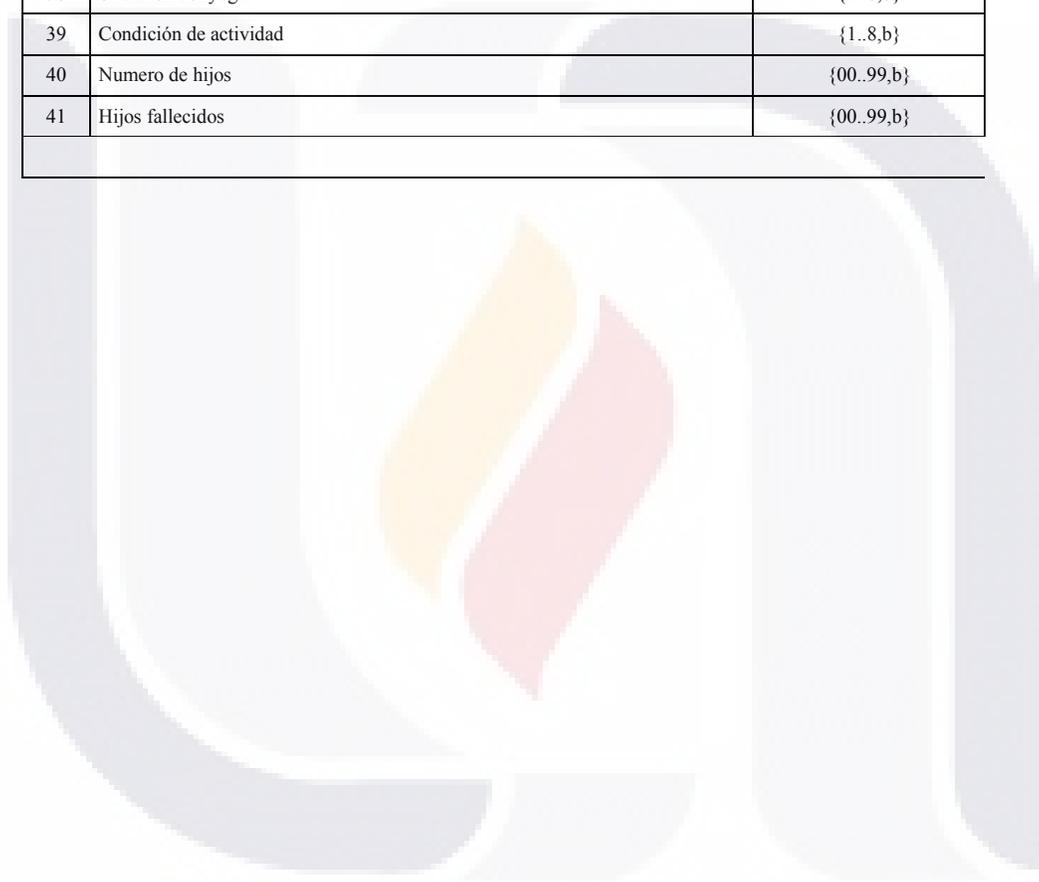
CENSO DE POBLACIÓN Y VIVIENDA 2010
VARIABLES DE SALIDA DE CAPTURA POR TABLA

(captura tradicional)

CAP##_TR_PERSONA

Cons.	Descripción	Rango Válido
1	Consecutivo de CM	{1..9999}
2	Consecutivo de persona en CM	{1..999999}
3	Consecutivo de cuestionario en CM	{1..99999}
4	Identificador único de vivienda	{1..99999999}
5	Número automático de persona en el cuestionario	{1..6}
6	Número automático de persona en la vivienda	{00000..99999}
7	Número de persona	{00000..99999}
8	Sexo	{1,3,b}
9	Edad	{000..999,b}
10	Parentesco (Jefa o jefe)	{1,3,b}
11	Parentesco	{1..9,b}
12	Entidad o país de nacimiento (aquí, en este estado)	{1,b}
13	Entidad o país de nacimiento (en otro estado)	{Alfanumérico,b}
14	Entidad o país de nacimiento (en los Estados Unidos de América)	{3,b}
15	Entidad o país de nacimiento (en otro país)	{Alfanumérico,b}
16	Derechohabencia a servicios de salud (primera institución)	{1..8,b}
17	Derechohabencia a servicios de salud (segunda institución)	{1..8,b}
18	Religión	{Alfanumérico,b}
19	Discapacidad (caminar, moverse, subir o bajar?)	{10,b}
20	Discapacidad (ver, aun usando lentes?)	{11,b}
21	Discapacidad (hablar, comunicarse o conversar?)	{12,b}
22	Discapacidad (oír, aun usando aparato auditivo?)	{13,b}
23	Discapacidad (vestirse, bañarse o comer?)	{14,b}
24	Discapacidad (poner atención o aprender cosas sencillas?)	{15,b}
25	Discapacidad (limitación mental?)	{16,b}
26	Discapacidad (no tiene)	{17,b}
27	Hablante de lengua indígena	{1,3,b}
28	Nombre de lengua indígena	{Alfanumérico,b}

29	Habla español	{1,3,b}
30	Asistencia	{1,3,b}
31	Escolaridad (grado)	{0..9,b}
32	Escolaridad (nivel académico)	{00..12,b}
33	Alfabetismo	{1,3,b}
34	Entidad o país de residencia en 2005 (aquí, en este estado)	{1,b}
35	Entidad o país de residencia en 2005 (en otro estado)	{Alfanumérico,b}
36	Entidad o país de residencia en 2005 (en los Estados Unidos de América)	{3,b}
37	Entidad o país de residencia en 2005 (en otro país)	{Alfanumérico,b}
38	Situación conyugal	{1..8,b}
39	Condición de actividad	{1..8,b}
40	Numero de hijos	{00..99,b}
41	Hijos fallecidos	{00..99,b}



CENSO DE POBLACIÓN Y VIVIENDA 2010

VARIABLES DE SALIDA DE CAPTURA POR TABLA

(captura tradicional)

CAP##_TR_PERSONA_AMP

Cons.	Descripción	Rango Válido
1	Consecutivo de CM	{1..9999}
2	Consecutivo de persona en CM	{1..999999}
3	Identificador único de vivienda	{1..99999999}
4	Parentesco (otro)	{Alfanumérico,b}
5	Identificación de la madre	{00..99,b}
6	Identificación de la madre (código)	{88,b}
7	Identificación del padre	{00..99,b}
8	Identificación del padre (código)	{88,b}
9	Uso de servicios de salud	{1..9,b}
10	Discapacidad (caminar, moverse, subir o bajar?)	{1..5,b}
11	Discapacidad (ver, aun usando lentes?)	{1..5,b}
12	Discapacidad (hablar, comunicarse o conversar?)	{1..5,b}
13	Discapacidad (oír, aun usando aparato auditivo?)	{1..5,b}
14	Discapacidad (vestirse, bañarse o comer?)	{1..5,b}
15	Discapacidad (poner atención o aprender cosas sencillas?)	{1..5,b}
16	Discapacidad (limitación mental?)	{1..5,b}
17	Comprensión de lengua indígena	{1,3,b}
18	Autoadscripción étnica	{1,3,b}
19	Antecedente escolar	{1..5,b}
20	Nombre de la carrera	{Alfanumérico,b}
21	Municipio de residencia en 2005	{1,b}
22	Municipio de residencia en 2005 (otro)	{Alfanumérico,b}
23	Identificación del cónyuge	{00..99,b}
24	Identificación del cónyuge	{88,b}
25	Verificación de condición de actividad	{1..7,b}
26	Ocupación u oficio (tareas)	{Alfanumérico,b}
27	Ocupación u oficio (puesto)	{Alfanumérico,b}
28	Posición en el trabajo	{1..6,b}

29	Prestaciones laborales (servicio médico)	{1,2,b}
30	Prestaciones laborales (aguinaldo)	{3,4,b}
31	Prestaciones laborales (vacaciones con goce de sueldo)	{1,2,b}
32	Prestaciones laborales (reparto de utilidades o prima vacacional)	{3,4,b}
33	Prestaciones laborales (ahorro para el retiro)	{1,2,b}
34	Prestaciones laborales (otras)	{3,4,b}
35	Horas trabajadas	{000..999,b}
36	Ingresos por trabajo (periodo)	{1..5,b}
37	Ingresos por trabajo (cantidad)	{000000..999999,b}
38	Unidad económica (donde trabajó)	{Alfanumérico,b}
39	Actividad económica (a que se dedica)	{Alfanumérico,b}
40	Municipio, entidad o país de trabajo (aquí en el municipio)	{1,b}
41	Municipio, entidad o país de trabajo (en otro municipio)	{Alfanumérico,b}
42	Municipio, entidad o país de trabajo (aquí en este estado)	{3,b}
43	Municipio, entidad o país de trabajo (en otro estado o país)	{Alfanumérico,b}
44	Otros ingresos (uno o más programas de gobierno?)	{1,2,b}
45	Otros ingresos (jubilación o pensión?)	{3,4,b}
46	Otros ingresos (ayuda de personas que viven en otro país?)	{1,2,b}
47	Otros ingresos (ayuda de personas que viven dentro del país?)	{3,4,b}
48	Otros ingresos (otras fuentes?)	{1,2,b}
49	Hijos sobrevivientes	{00..99,b}
50	Fecha de nacimiento (mes)	{00..99,b}
51	Fecha de nacimiento (año)	{0000..9999,b}
52	Sobrevivencia	{1,3,b}
53	Edad al morir (días)	{00..99,b}
54	Edad al morir (meses)	{00..99,b}
55	Edad al morir (años)	{00..99,b}

CENSO DE POBLACIÓN Y VIVIENDA 2010

VARIABLES DE SALIDA DE CAPTURA POR TABLA

(captura tradicional)

CAP##_TR_MIGINT_AMP

Cons.	Descripción	Rango Válido
1	Consecutivo de CM	{1..9999}
2	Consecutivo de migrante en CM	{1..9999}
3	Consecutivo de cuestionario en CM	{1..99999}
4	Identificador único de vivienda	{1..99999999}
5	Número automático de migrante	{1..4}
6	Número de migrante	{00..99}
7	Condición de residencia	{1,3,b}
8	Sexo	{1,3,b}
9	Edad	{000..999,b}
10	Fecha de emigración (mes)	{00..99,b}
11	Fecha de emigración (año)	{0000..9999,b}
12	Lugar de origen (estado)	{Alfanumérico,b}
13	País de destino (Estados Unidos de América)	{1,b}
14	País de destino (otro país)	{Alfanumérico,b}
15	País de residencia	{1..3,b}
16	Fecha de retorno (mes)	{00..99,b}
17	Fecha de retorno (año)	{0000..9999,b}
18	Condición de residencia actual	{1,3,b}
19	Número del migrante en la lista de personas	{00..99,b}

CENSO DE POBLACIÓN Y VIVIENDA 2010

VARIABLES DE SALIDA DE CAPTURA POR TABLA

(captura tradicional)

CAP##_TR_CUADERNILLO_POS

Cons.	Descripción	Rango Válido
1	Consecutivo de CM	{1..9999}
2	Identificación de cuadernillo en CM	{1..9999}
3	Identificación de paquete en CM	{1..9999}
4	Entidad federativa	{01..32}
5	Municipio o delegación	{001..570}
6	AGEB	{001..999;0..9 ó A,P}
7	Localidad	{0001..9999}
8	Manzana	{000..999}
9	Segmento	{A..I,N,P,Q,V}
10	Número de cuadernillo	{00..99}
11	Total de cuadernillos	{00..99}
12	Total de cédulas utilizadas en el cuadernillo	{00..99,b}
13	Folio de paquete	{101..999}
14	Consecutivo del cuadernillo en el paquete	{00..99}
15	Clave del posenumerador(a)	{0..9,b}
16	Clave de supervisor de posenumeradores	{0..9,b}

CENSO DE POBLACIÓN Y VIVIENDA 2010

VARIABLES DE SALIDA DE CAPTURA POR TABLA

(captura tradicional)

CAP##_TR_VIVIENDA_POS

Cons.	Descripción	Rango Válido
1	Consecutivo de CM	{1..9999}
2	Consecutivo de cédula en CM	{1..99999}
3	Identificación de cuademillo en CM	{1..9999}
4	Identificador único de vivienda	{1..99999999}
5	Apellido	{0..9}
6	Consecutivo de la vivienda	{000..999,b}
7	Número de cédula	{01..99}
8	Total de cédulas en la vivienda	{01..99}
9	Fecha de levantamiento (Día)	{00..99,b}
10	Fecha de levantamiento (Mes)	{00..99,b}
11	Etiqueta	{1..5,b}
12	Consecutivo de la vivienda enumeración	{000..999,b}
13	Total de cuestionarios de la vivienda enumeración	{00..99,b}
14	Tipo de cotejo	{1,2,b}
15	Número del informante	{00..99,b}
16	Verificación de la lista	{1,3,b}
17	Salida de residentes	{1,3,b}
18	Bandera para liberar documentos sin información	{0,1}

CENSO DE POBLACIÓN Y VIVIENDA 2010

VARIABLES DE SALIDA DE CAPTURA POR TABLA

(captura tradicional)

CAP##_TR_PERSONA_POS

Cons.	Descripción	Rango Válido
1	Consecutivo de CM	{1..9999}
2	Identificación de persona en CM	{1..999999}
3	Consecutivo de cédula en CM	{1..99999}
4	Identificador único de vivienda	{1..99999999}
5	Tipo de residente	{1,2}
6	Número automático de persona	{1..6}
7	Código de cotejo	{1..6,b}
8	Número de renglón	{00..99,b}
9	Sexo	{1,3,b}
10	Edad	{000..999,b}
11	Parentesco	{1..9,b}
12	Condición de residencia	{1,3,b}
13	Fecha de llegada (Día)	{00..99,b}
14	Fecha de llegada (Mes)	{00..99,b}
15	Motivo de no residencia	{1..3,b}
16	Fecha de salida (Día)	{00..99,b}
17	Fecha de salida (Mes)	{00..99,b}

CENSO DE POBLACIÓN Y VIVIENDA 2010
VARIABLES DE SALIDA DE CAPTURA POR TABLA

(captura tradicional)

CAP##_TR_CUESTIONARIO_LOC

Cons.	Descripción	Rango Válido
1	Consecutivo de CM	{1..9999}
2	Consecutivo de cuestionario en CM	{1..9999}
3	Consecutivo de paquete en CM	{1..9999}
4	Entidad federativa	{01..32}
5	Municipio o delegación	{001..570}
6	AGEB	{000..999;0..9 ó A,P}
7	Localidad	{0001..9999}
8	Apellido	{0..9}
9	Folio de paquete	{101..999}
10	Consecutivo del cuestionario en el paquete	{001..999}
11	Clave de supervisor	{0..9,b}
12	Clave de validador	{0..9,b}
13	Fecha del levantamiento (día)	{00..99,b}
14	Fecha del levantamiento (mes)	{00..99,b}
15	Localidad señalada para llenado	{1,3,b}
16	Situación de la localidad	{1,3,b}
17	Total de inmuebles estimados (en ruinas)	{00..99,b}
18	Total de inmuebles estimados (en construcción)	{00..99,b}
19	Total de inmuebles estimados (local con actividad)	{00..99,b}
20	Total de inmuebles estimados (local vacío)	{00..99,b}
21	Total de inmuebles estimados (vivienda)	{00..99,b}
22	Total de inmuebles estimados (otra)	{00..99,b}
23	Tipo y número de inmuebles (descripción otra)	{Alfanumérico,b}
24	Nombre	{Alfanumérico,b}
25	Sexo	{1,3,b}
26	Edad	{00..99,b}
27	Puesto	{1..5,b}
28	Puesto (Otro)	{Alfanumérico,b}

29	Tiempo de residencia	{00..99,b}
30	Escolaridad (grado)	{0..9,b}
31	Escolaridad (nivel académico)	{0..4,b}
32	Bandera para liberar documentos sin información	{0,1}
33	Carretera	{1..3,b}
34	Distancia a carretera	{000..999,b}
35	Tiempo de traslado (Horas)	{00..99,b}
36	Tiempo de traslado (Minutos)	{00..99,b}
37	Tiempo de traslado (cabecera municipal o no hay transporte)	{88,89,b}
38	Transporte público	{1..6,b}
39	Frecuencia	{00..99,b}
40	Costo	{000..999,b}
41	Red de agua potable	{1,3,b}
42	Cobro de agua	{1,3,b}
43	Red de drenaje	{1,3,b}
44	Destino del drenaje	{1..4,b}
45	Limpieza y recolección de basura (limpieza de calles)	{1,2,b}
46	Limpieza y recolección de basura (recolección de basura de las viviendas)	{3,4,b}
47	Destino de la basura	{1..5,b}
48	Equipamiento (alumbrado público)	{1,2,b}
49	Equipamiento (calles pavimentadas)	{3,4,b}
50	Equipamiento (pozo de agua comunitario)	{1,2,b}
51	Equipamiento (plaza o jardín)	{3,4,b}
52	Equipamiento (panteón)	{1,2,b}
53	Servicios públicos (oficina o agencia municipal?)	{1,2,b}
54	Servicios públicos (oficina de registro civil?)	{3,4,b}
55	Servicios públicos (policía preventiva?)	{1,2,b}
56	Servicios públicos (cárcel o calabozo?)	{3,4,b}
57	Telecomunicaciones (teléfono público?)	{1,2,b}
58	Telecomunicaciones (caseta de teléfono de larga distancia?)	{3,4,b}
59	Telecomunicaciones (internet público?)	{1,2,b}
60	Comercio (mercado fijo?)	{1,2,b}
61	Comercio (tianguis?)	{3,4,b}
62	Comercio (tienda de abarrotes?)	{1,2,b}
63	Comercio (farmacia?)	{3,4,b}
64	Comercio (papelería?)	{1,2,b}
65	Comercio (tienda o ropa de calzado?)	{3,4,b}

66	Comercio (tienda de muebles o aparatos electrodomésticos?)	{1,2,b}
67	Comercio (tienda de materiales de construcción?)	{3,4,b}
68	Comercio (venta de gas?)	{1,2,b}
69	Abasto de alimentos (frijol?)	{1,2,b}
70	Abasto de alimentos (maíz o harina de maíz)	{3,4,b}
71	Abasto de alimentos (arroz?)	{1,2,b}
72	Abasto de alimentos (leche?)	{3,4,b}
73	Abasto de alimentos (carne de res o de puerco?)	{1,2,b}
74	Abasto de alimentos (pollo?)	{3,4,b}
75	Abasto de alimentos (pescado fresco?)	{1,2,b}
76	Servicios comunales (oficina para enviar y recibir dinero?)	{1,2,b}
77	Servicios comunales (casa de huéspedes?)	{3,4,b}
78	Servicios comunales (hotel o motel?)	{1,2,b}
79	Servicios comunales (fonda?)	{3,4,b}
80	Servicios comunales (restaurante?)	{1,2,b}
81	Servicios comunales (cantina, bar o depósito de licores?)	{3,4,b}
82	Salud (Clínica o centro de salud?)	{1,2,b}
83	Salud (consultorio médico particular?)	{3,4,b}
84	Salud (partera o comadrona?)	{1,2,b}
85	Salud (curandera(o)?)	{3,4,b}
86	Escuelas (preescolar?)	{1,2,b}
87	Escuelas (primaria completa?)	{3,4,b}
88	Escuelas (primaria incompleta?)	{1,2,b}
89	Escuelas (telesecundaria o secundaria?)	{3,4,b}
90	Escuelas (preparatoria o bachillerato?)	{1,2,b}
91	Recintos culturales y deportivos (biblioteca?)	{1,2,b}
92	Recintos culturales y deportivos (casa de la cultura?)	{3,4,b}
93	Recintos culturales y deportivos (cancha deportiva?)	{1,2,b}
94	Lengua indígena	{1,3,b}
95	Nombre de la lengua indígena	{Alfanumérico,b}
96	Lugar de uso (la escuela?)	{1,2,b}
97	Lugar de uso (la iglesia?)	{3,4,b}
98	Lugar de uso (las festividades?)	{1,2,b}
99	Lugar de uso (el hogar?)	{3,4,b}
100	Lugar de uso (para comprar o vender?)	{1,2,b}
101	Actividades económicas 1	{Alfanumérico,b}
102	Actividades económicas 2	{Alfanumérico,b}

103	Actividades económicas 3	{Alfanumérico,b}
104	Conflictos (la propiedad de la tierra?)	{1,2,b}
105	Conflictos (preferencias religiosas?)	{3,4,b}
106	Conflictos (preferencias electorales?)	{1,2,b}
107	Conflictos (delincuencia?)	{3,4,b}
108	Conflictos (alcoholismo o drogadicción?)	{1,2,b}
109	Problema principal	{Alfanumérico,b}



CENSO DE POBLACIÓN Y VIVIENDA 2010

VARIABLES DE SALIDA DE CAPTURA POR TABLA

(captura tradicional)

CAP##_TR_CUESTIONARIO_URB

Cons.	Descripción	Rango Válido
1	Consecutivo de CM	{1..9999}
2	Consecutivo del cuestionario en CM	{1..9999}
3	Consecutivo de paquete en CM	{1..9999}
4	Entidad federativa	{01..32}
5	Municipio o delegación	{001..570}
6	AGEB	{001..999;0..9 ó A,P}
7	Localidad	{0001..9999}
8	Manzana	{001..999}
9	Apellido	{0..9}
10	Número de cuestionario en la manzana	{01..99}
11	Total de cuestionarios en la manzana	{01..99}
12	Consecutivo del cuestionario en el paquete	{001..999}
13	Fecha de recorrido (día)	{00..99;b}
14	Fecha de recorrido (mes)	{00..99;b}
15	Clave de supervisor(a)	{0..9;b}
16	Clave de validador(a)	{0..9;b}
17	Manzana de conjunto habitacional	{1,3,b}
18	Acceso a la manzana (toda la manzana)	{1,2,b}
19	Acceso a la manzana (alguna vialidad)	{3,4,b}
20	Acceso a la manzana (edificios o conjuntos habitacionales)	{1,2,b}
21	Precauciones (personas que obstaculizan)	{1,2,b}
22	Precauciones (personas ebrias o drogadas)	{3,4,b}
23	Precauciones (vigilancia)	{1,2,b}
24	Bandera para liberar documentos sin información	{0,1}

CENSO DE POBLACIÓN Y VIVIENDA 2010
VARIABLES DE SALIDA DE CAPTURA POR TABLA

(captura tradicional)

CAP##_TR_VIALIDAD_URB

Cons.	Descripción	Rango Válido
1	Consecutivo de CM	{1..9999}
2	Consecutivo de vialidad en CM	{1..99999}
3	Consecutivo del cuestionario en CM	{1..9999}
4	Número automático de la vialidad	{1..6}
5	Nombre de la vialidad	{Alfanumérico,b}
6	Tipo de vialidad o rasgo	{1..4,b}
7	Acceso a personas	{1..3,b}
8	Acceso de automóviles	{1,2,b}
9	Recubrimiento de la calle	{1..3,b}
10	Letrero con nombre de calle (señalización)	{1..4,b}
11	Alumbrado público	{1..4,b}
12	Teléfono público	{1..4,b}
13	Alcantarillas de drenaje pluvial	{1..4,b}
14	Banqueta	{1..4,b}
15	Guarnición	{1..4,b}
16	Árboles o plantas de ornato	{1..4,b}
17	Rampas para silla de ruedas	{1..4,b}
18	Transporte colectivo	{1..4,b}
19	Puestos de comercio semifijo	{1..4,b}
20	Puestos de comercio ambulante	{1..4,b}

Descripción de los procesos tradicionales para la recopilación, integración y registro de la información censal.

En las coordinaciones municipales se encuentra el responsable de procesos de producción quien se encarga de administrar y configurar las aplicaciones que se utilizan CM, asigna cargas de trabajo de captura, asigna cargas de trabajo de verificación, integra la información de los capturistas, realiza los envíos WEB

(diarios) de la información recabada al termino del turno de captura, lanza los procesos primarios y genera los reportes de avance y seguimiento a la captura. Mientras tanto el técnico de zona con funciones específicas apoya el trabajo del responsable de procesos de producción monitoreando los procesos de envío y descargas de insumos para las aplicaciones y los capturistas realizan el registro de todos los instrumentos captación además de la verificación de la captura al 100% excepto el cuestionario básico.

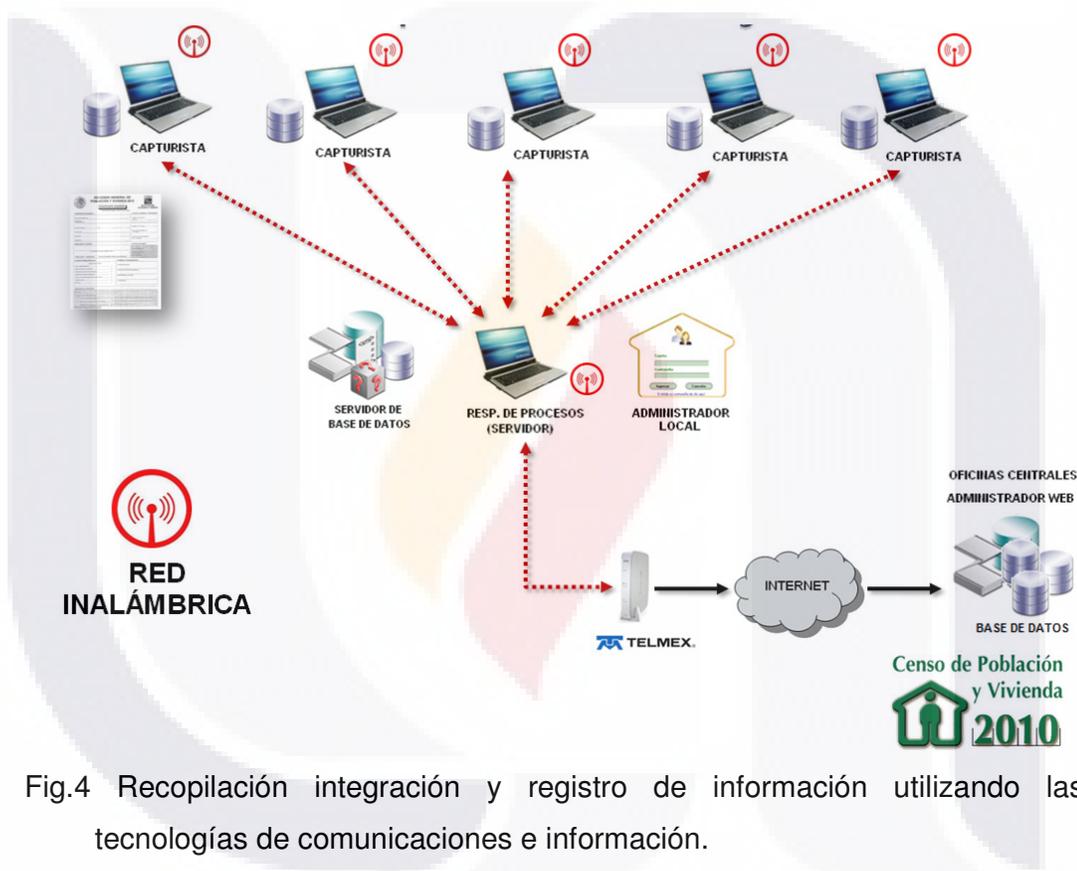


Fig.4 Recopilación integración y registro de información utilizando las tecnologías de comunicaciones e información.

Definición de entidades y atributos

Las entidades y atributos a que se refiere este tema son aquellos sobre los que las siguientes etapas del procesamiento de la información trabajan, es decir, el resultado de la transformación que tienen los datos capturados y mostrados en el tema “Descripción de los datos proporcionados por cada fuente de información” y de los que ya se mostró el modelo de datos en la fig. 3.

Las siguientes tablas muestran la descripción de dichas entidades y atributos producto de la transformación.

CENSO DE POBLACIÓN Y VIVIENDA 2010
ESTRUCTURA SEGMENTOS
 LOCALIDADES, MANZANAS Y/O SEGMENTOS

TR_SEGMENTO

Cons.	Descripción	Rango Válido
1	Identificación de segmento	{1..9999999}
2	Entidad Federativa	{01..32}
3	Municipio o Delegación	{001..570}
4	Clave de AGEB	{001..999; 0..9 ó A,P}
5	Localidad	{0001..9999}
6	Manzana	{000..999}
7	Segmento	{A..I,N,P,Q,R,S,T,W,Y}
8	Identificación de Coordinación Estatal	{01..34}
9	Identificación de lote	{01..34;001..570; 01..31;06..10}

CENSO DE POBLACIÓN Y VIVIENDA 2010
ESTRUCTURA CARACTERÍSTICAS DE LAS VIVIENDAS
VARIABLES COMUNES BÁSICO Y AMPLIADO

TR_VIVIENDA

Cons.	Descripción	Rango Válido
1	Identificación de vivienda	{1..99999999}
2	Identificación de segmento	{1..9999999}
3	Tipo de documento	{1..3}
4	Medio de captación (aplicación de captura)	{1..7}
5	Estado de codificación (registro codificado y/o afectado)	{0..1}
6	Consecutivo de la vivienda	{001..999}
7	Apellido	{0..9}
8	Clase de vivienda particular	{1..7,b}
9	Clase de vivienda colectiva	{Alfanumérico,b}
10	Clase de vivienda colectiva (Clave)	{01..16,b}
11	Clase de vivienda colectiva (Indicador)	{00..07,'YY','MM'}
12	Piso	{1..3,b}
13	Cuartos dormitorio	{00..99,b}
14	Total de cuartos	{00..99,b}
15	Electricidad	{1,3,b}
16	Agua entubada	{1..6,b}
17	Excusado	{1,3,b}
18	Admisión de agua	{5..7,b}
19	Drenaje	{1..5,b}
20	Electrodomésticos (aparato con radio)	{1,2,b}
21	Electrodomésticos (televisor)	{3,4,b}
22	Electrodomésticos (refrigerador)	{1,2,b}
23	Electrodomésticos (lavadora)	{3,4,b}
24	Electrodomésticos (automóvil o camioneta)	{1,2,b}
25	Electrodomésticos (computadora)	{3,4,b}

26	Electrodomésticos (línea telefónica fija)	{1,2,b}
27	electrodomésticos (teléfono celular)	{3,4,b}
28	Electrodomésticos (internet)	{1,2,b}
29	Número del informante	{00..99,b}
30	Número de personas en la vivienda	{00000..99999,b}
31	Verificación de la lista	{1,3,b}



CENSO DE POBLACIÓN Y VIVIENDA 2010
ESTRUCTURA CARACTERÍSTICAS DE LAS VIVIENDAS
VARIABLES EXCLUSIVAS DE AMPLIADO

TR_VIVIENDA_AMP

Cons.	Descripción	Rango Válido
1	Identificación de vivienda	{1..99999999}
2	Estado de codificación (registro codificado y/o afectado)	{0..1}
3	Paredes	{1..8,b}
4	Techos	{1..9,b}
5	Cocina	{1,3,b}
6	Dotación de agua	{1..5,b}
7	Uso exclusivo	{1,3,b}
8	Combustible	{1..6,b}
9	Eliminación de basura	{1..8,b}
10	Tenencia	{1..3,b}
11	Adquisición	{1..4,b}
12	Equipamiento (estufa de gas)	{1,2,b}
13	Equipamiento (estufa de leña o carbón con chimenea)	{3,4,b}
14	Equipamiento (tinaco)	{1,2,b}
15	Equipamiento (calentador de agua)	{3,4,b}
16	Equipamiento (cisterna o aljibe)	{1,2,b}
17	Equipamiento (regadera)	{3,4,b}
18	Equipamiento (medidor de luz)	{1,2,b}
19	Condición de migración internacional	{1,3,b}
20	Número de migrantes internacionales	{00..99,b}
21	Acceso a la alimentación (¿alguna persona... sólo comió una vez al día?)	{1,2,b}
22	Acceso a la alimentación (¿alguna persona... dejó de comer todo un día?)	{3,4,b}
23	Acceso a la alimentación (¿alguna vez se quedaron sin comida?)	{1,2,b}

CENSO DE POBLACIÓN Y VIVIENDA 2010
ESTRUCTURA CARÁTULA DE CUESTIONARIOS
VARIABLES COMUNES BÁSICO Y AMPLIADO

TR_CUESTIONARIO

Cons.	Descripción	Rango Válido
1	Identificación de cuestionario	{1..99999999}
2	Identificación de vivienda	{1..99999999}
3	Identificación de paquete	{1..99999999}
4	Estado de codificación (registro codificado y/o afectado)	{0..1}
5	Número de cuestionario en la vivienda	{0001..9999}
6	Total de cuestionarios en la vivienda	{0001..9999}
7	Consecutivo del cuestionario en el paquete	{001..999}
8	Fecha de entrevista (día)	{01..31,b}
9	Fecha de entrevista (mes)	{05..07,b}
10	Clave de entrevistador(a)	{1..9,b}
11	Clave de supervisor(a)	{1..9,b}
12	Clave de validador(a)	{1..9,b}
13	Clave de verificador(a)	{1..9,b}
14	Total de ocupantes	{00..99,b}

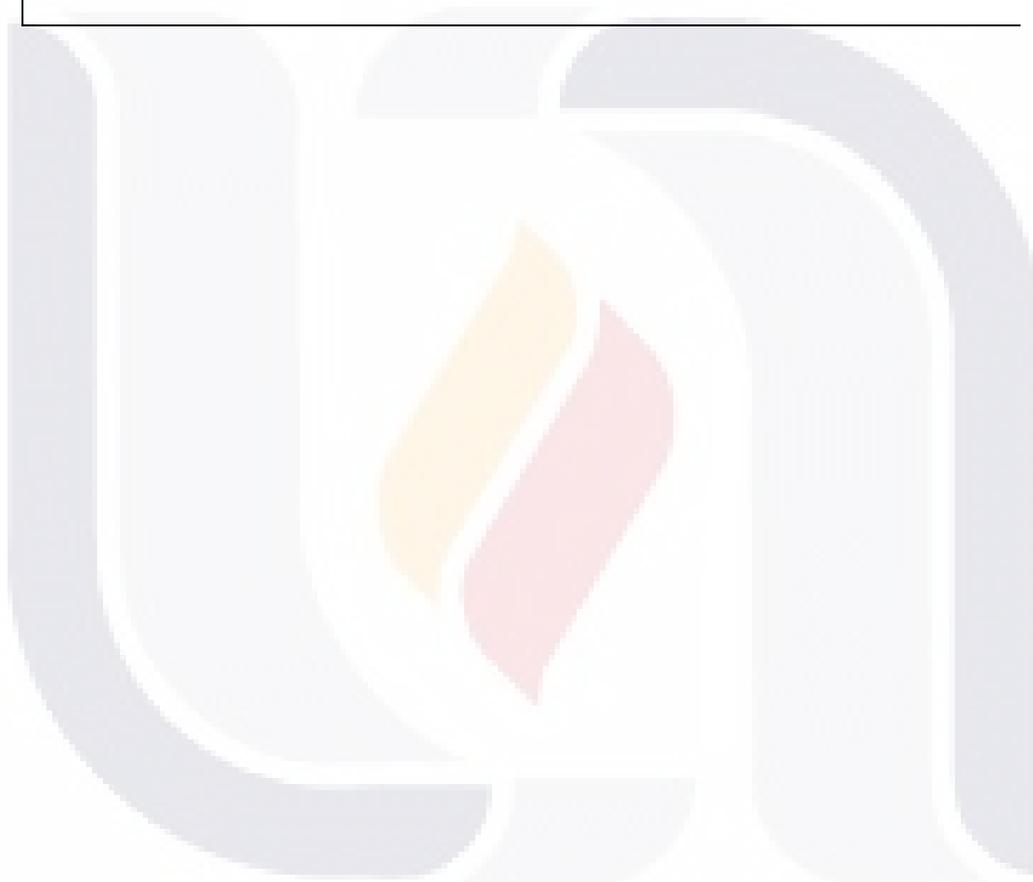
CENSO DE POBLACIÓN Y VIVIENDA 2010
ESTRUCTURA CARACTERÍSTICAS DE LAS PERSONAS
VARIABLES COMUNES BÁSICO Y AMPLIADO

TR_PERSONA

Cons.	Descripción	Rango Válido
1	Identificación de persona	{1..999999999}
2	Entidad Federativa	{01..32}
3	Municipio o Delegación	{001..570}
4	Identificación de vivienda	{1..999999999}
5	Identificación de cuestionario	{1..999999999}
6	Estado de codificación (registro codificado y/o afectado)	{0..1}
7	Número automático de persona	{1..6}
8	Número automático de persona en la vivienda	{00001..99999}
9	Número de persona	{00000..99999}
10	Sexo	{1,3,b}
11	Edad	{000..999,b}
12	Jefa o Jefe (autoenumeración)	{1,3,b}
13	Parentesco	{1..9,b}
14	Entidad o país de nacimiento (aquí, en este estado)	{1,b}
15	Entidad o país de nacimiento (en otro estado)	{Alfanumérico,b}
16	Entidad o país de nacimiento (en otro estado -clave-)	{001-032, 100-535, 600, 800, 888, 999, b}
17	Entidad o país de nacimiento (en otro estado -indicador-)	{00..04,10..12,'YY','MM'}
18	Entidad o país de nacimiento (en los Estados Unidos de América)	{3,b}
19	Entidad o país de nacimiento (en otro país)	{Alfanumérico,b}
20	Entidad o país de nacimiento (en otro país -clave-)	{001-032, 100-535, 600, 800, 888, 999, b}
21	Entidad o país de nacimiento (en otro país -indicador-)	{00..04,10..12,'YY','MM'}
22	Derechohabencia a servicios de salud (primera institución)	{1..8,b}
23	Derechohabencia a servicios de salud (segunda institución)	{1..8,b}
24	Religión	{Alfanumérico,b}

25	Religión (Clave)	{110100-110200, 120100-120600, 130100-139999, 140100-140303, 150100-150300, 210100-210900, 220100-220200, 230100-230500, 240100-241200, 250100-250800, 260100-261000, 270100-271300, 280100-289999, 290100-290600, 310000, 310100, 888888, 888899, 999999, b}
26	Religión (indicador)	{00-08, 'YY', 'MM', b}
27	Discapacidad (caminar, moverse, subir o bajar?)	{10,b}
28	Discapacidad (ver, aun usando lentes?)	{11,b}
29	Discapacidad (hablar, comunicarse o conversar?)	{12,b}
30	Discapacidad (oír, aun usando aparato auditivo?)	{13,b}
31	Discapacidad (vestirse, bañarse o comer?)	{14,b}
32	Discapacidad (poner atención o aprender cosas sencillas?)	{15,b}
33	Discapacidad (limitación mental?)	{16,b}
34	Discapacidad (no tiene)	{17,b}
35	Hablante de lengua indígena	{1,3,b}
36	Nombre de lengua indígena	{Alfanumérico,b}
37	Nombre de lengua indígena (Clave)	{0111-0131, 0200-0243, 0311-0342, 0400-0494, 0511, 0611, 0711,0712, 0800-0824, 0911-0981, 1011-1041, 1111, 1211,1311, 5001-5136, 5501-5602, 6000, 7000, 8000, 9999,b}
38	Nombre de lengua indígena (Indicador)	{00-08,'YY','MM',b}
39	Hablante también español	{1,3,b}
40	Asistencia	{1,3,b}
41	Escolaridad (grado)	{0..9,b}
42	Escolaridad (nivel académico)	{00..12,b}
43	Alfabetismo	{1,3,b}
44	Entidad o país de residencia en 2005 (aquí, en este estado)	{1,b}
45	Entidad o país de residencia en 2005 (en otro estado)	{Alfanumérico,b}
46	Entidad o país de residencia en 2005 (en otro estado -Clave-)	{001-032, 100-535, 600, 800, 888, 999, b}
47	Entidad o país de residencia en 2005 (en otro estado -Indicador-)	{01-04,10- 12,'YY','MM',b}

48	Entidad o país de residencia en 2005 (en los EUA)	{3,b}
49	Entidad o país de residencia en 2005 (en otro país)	{Alfanumérico,b}
50	Entidad o país de residencia en 2005 (en otro país -Clave-)	{001-032, 100-535, 600, 800, 888, 999, b}
51	Entidad o país de residencia en 2005 (en otro país -Indicador-)	{01-04,10-12,'YY','MM',b}
52	Situación conyugal	{1..8,b}
53	Condición de actividad	{1..8,b}
54	Número de hijos	{00..99,b}
55	Hijos fallecidos	{00..99,b}



CENSO DE POBLACIÓN Y VIVIENDA 2010
ESTRUCTURA CARACTERÍSTICAS DE LAS PERSONAS
 VARIABLES EXCLUSIVAS DE AMPLIADO

TR_PERSONA_AMP

Cons.	Descripción	Rango Válido
LLAVE ÚNICA DE PERSONA		
1	Identificación de persona	{1..999999999}
IDENTIFICACIÓN DEL PAQUETE		
2	Entidad Federativa	{01..32}
3	Municipio o Delegación	{001..570}
VARIABLES AUXILIARES		
4	Estado de codificación (registro codificado y/o afectado)	{0,1}
CARACTERÍSTICAS DE LAS PERSONAS		
5	Parentesco (otro)	{Alfanumérico,b}
6	Parentesco (otro -Clave-)	{101, 201-204, 301-305,401-413, 421, 431, 441, 451, 461,501-503, 601- 623,701, 711-715, 999,b}
7	Parentesco (otro -Indicador-)	{00-08,'YY','MM',b}
8	Identificación de la madre	{00..99,b}
9	Identificación de la madre (código)	{88,b}
10	Identificación del padre	{00..99,b}
11	Identificación del padre (código)	{88,b}
12	Uso de servicios de salud	{1..9,b}
13	Discapacidad (caminar, moverse, subir o bajar?)	{1..5,b}
14	Discapacidad (ver, aun usando lentes?)	{1..5,b}
15	Discapacidad (hablar, comunicarse o conversar?)	{1..5,b}
16	Discapacidad (oír, aun usando aparato auditivo?)	{1..5,b}
17	Discapacidad (vestirse, bañarse o comer?)	{1..5,b}
18	Discapacidad (poner atención o aprender cosas sencillas?)	{1..5,b}
19	Discapacidad (limitación mental?)	{1..5,b}
20	Comprensión de lengua indígena	{1,3,b}
21	Autoadscripción étnica	{1,3,b}
22	Antecedente escolar	{1..5,b}
23	Nombre de la carrera	{Alfanumérico,b}

24	Nombre de la carrera (Clave)	{11000-12703,19900, 21000-22500,29900, 31000-34200,39900, 41100-44100,49900, 51000-53200,59900, 61000-62100,69900, 71000-71700,79900, 81100-83300,89900, 80000, 99999,b}
25	Nombre de la carrera (Indicador)	{00-08,'YY','MM',b}
26	Municipio de residencia en 2005	{1,b}
27	Municipio de residencia en 2005 (otro)	{Alfanumérico,b}
28	Municipio de residencia en 2005 (otro -Clave-)	{001-570, 800,888,999,b}
29	Municipio de residencia en 2005 (otro -Indicador-)	{00-04, 10-12,'YY','MM',b}
30	Identificación del cónyuge	{00..99,b}
31	Identificación del cónyuge (código)	{88,b}
32	Verificación de condición de actividad	{1..7,b}
33	Ocupación u oficio (tareas)	{Alfanumérico,b}
34	Ocupación u oficio (tareas -Clave-)	{11110-17233, 19990-19992, 21110-28270, 29910, 29920, 31010-32322, 39990, 41110, 42010-42330, 43110-43122, 49990, 51010-54130, 59990, 61010-63112, 69990, 71010-76142, 79990, 81010-83522, 89990, 91110-97330, 98880, 98888, 99999, 10001-10008, b}
35	Ocupación u oficio (tareas -Indicador-)	{00-08,'YY','MM',b}
36	Ocupación u oficio (puesto)	{Alfanumérico,b}
37	Posición en el trabajo	{1..6,b}
38	Prestaciones laborales (servicio médico)	{1,2,b}
39	Prestaciones laborales (aguinaldo)	{3,4,b}
40	Prestaciones laborales (vacaciones con goce de sueldo)	{1,2,b}
41	Prestaciones laborales (reparto de utilidades o prima vacacional)	{3,4,b}
42	Prestaciones laborales (ahorro para el retiro)	{1,2,b}
43	Prestaciones laborales (otras)	{3,4,b}
44	Horas trabajadas	{000..999,b}
45	Ingresos por trabajo (periodo)	{1..5,b}
46	Ingresos por trabajo (pesos)	{000000..999999,b}
47	Unidad económica (donde trabajó)	{Alfanumérico,b}

48	Unidad económica (donde trabajó -Clave-)	{1110-1150,1199, 2110-2132,2199, 2210-2222, 2361-2382,2399, 3110-3160, 3210-3270, 3310-3380,3399, 4310-4370, 4399, 4611-4690,4699, 4810-4882, 4899,4910-4930, 5110-5190, 5199, 5210-5240, 5299, 5310-5330, 5399, 5411-5414, 5510, 5611-5620, 6111-6150,6199, 6211-6259,6299, 7111-7133, 7210-7223, 8111-8140, 9311-9320, 9399, 9888, 9999, 1001-1008, b}
49	Unidad económica (donde trabajó -Indicador-)	{00-08, 'YY','MM',b}
50	Actividad económica (a que se dedica)	{Alfanumérico,b}
51	Municipio, entidad o país de trabajo (aquí en el municipio)	{1,b}
52	Municipio, entidad o país de trabajo (en otro municipio)	{Alfanumérico,b}
53	Municipio, entidad o país de trabajo (en otro municipio -Clave-)	{001-570, 800,888,999,b}
54	Municipio, entidad o país de trabajo (en otro municipio -Indicador-)	{00-04, 10-12,'YY','MM',b}
55	Municipio, entidad o país de trabajo (aquí en este estado)	{3,b}
56	Municipio, entidad o país de trabajo (en otro estado o país)	{Alfanumérico,b}
57	Municipio, entidad o país de trabajo (en otro estado o país -Clave-)	{001-032,100-535, 600,800,888,999,b}
58	Municipio, entidad o país de trabajo (en otro estado o país -Indicador-)	{00-04, 10-12,'YY','MM',b}
59	Otros ingresos (uno o más programas de gobierno?)	{1,2,b}
60	Otros ingresos (jubilación o pensión?)	{3,4,b}
61	Otros ingresos (ayuda de personas que viven en otro país?)	{1,2,b}
62	Otros ingresos (ayuda de personas que viven dentro del país ?)	{3,4,b}
63	Otros ingresos (otras fuentes ?)	{1,2,b}
64	Hijos sobrevivientes	{00..99,b}
65	Fecha de nacimiento (mes)	{00..99,b}
66	Fecha de nacimiento (año)	{0000..9999,b}
67	Sobrevivencia	{1,3,b}
68	Edad al morir (días)	{00..99,b}
69	Edad al morir (meses)	{00..99,b}
70	Edad al morir (años)	{00..99,b}

CENSO DE POBLACIÓN Y VIVIENDA 2010
ESTRUCTURA MIGRACIÓN INTERNACIONAL
VARIABLES EXCLUSIVAS DE AMPLIADO

TR_MIGINT_AMP

Cons.	Descripción	Rango Válido
1	Identificación de migrante internacional	{1..9999999}
2	Identificación de vivienda	{1..99999999}
3	Identificación de cuestionario	{1..99999999}
4	Estado de codificación (registro codificado y/o afectado)	{0,1}
5	Número automático de migrante	{1..4}
6	Número de migrante	{00..99}
7	Condición de residencia	{1,3,b}
8	Sexo	{1,3,b}
9	Edad	{000..999,b}
10	Fecha de emigración (mes)	{00..99,b}
11	Fecha de emigración (año)	{0000..9999,b}
12	Lugar de origen (estado)	{Alfanumérico,b}
13	Lugar de origen (estado -Clave-)	{001-032, 800, 999, b}
14	Lugar de origen (estado -Indicador-)	{00-04, 10-12,'YY','MM',b}
15	País de destino (Estados Unidos de América)	{1,b}
16	País de destino (otro país)	{Alfanumérico,b}
17	País de destino (otro país -Clave-)	{100-170, 200-250, 300-358, 400-453, 500-535, 600, 700, 800, 999, b}
18	País de destino (otro país -Indicador-)	{00-04, 10-12,'YY','MM',b}
19	País de residencia	{1..3,b}
20	Fecha de retorno (mes)	{00..99,b}
21	Fecha de retorno (año)	{0000..9999,b}
22	Condición de residencia actual	{1,3,b}
23	Número del migrante en la lista de personas	{00..99,b}

CENSO DE POBLACIÓN Y VIVIENDA 2010
ESTRUCTURA LISTADO DE INMUEBLES
VARIABLES DE CARÁTULA Y RESUMEN

TR_LISTADO

Cons.	Descripción	Rango Válido
1	Identificación de listado	{1..9999999}
2	Identificación de segmento	{1..9999999}
3	Apellido	{0..9}
4	Número de listado	{01..99}
5	Total de listados en la manzana, localidad o segmento	{01..99}
6	Tipo de cuestionario	{1..3,b}
7	Clave de entrevistador(a)	{0..9,b}
8	Clave de supervisor(a)	{0..9,b}
9	Clave de validador(a)	{0..9,b}
10	Clave de verificador(a)	{0..9,b}
11	Fecha de inicio de recorrido (día)	{01..31,b}
12	Fecha de inicio de recorrido (mes)	{05..07,b}
13	Fecha de término de recorrido (día)	{01..31,b}
14	Fecha de término de recorrido (mes)	{05..07,b}
15	Manzana o localidad	{1..3,b}
16	Tipo de documento	{1..4}
17	Bandera para liberar documentos sin información	{0,1}
18	Viviendas particulares con entrevista completa	{000..999,b}
19	Viviendas particulares con entrevista incompleta	{000..999,b}
20	Viv. part. con ausencia de residentes al momento de la vista	{000..999,b}
21	Viviendas particulares donde se negaron a dar información	{000..999,b}
22	Viviendas particulares deshabitadas	{000..999,b}
23	Viviendas particulares de uso temporal	{000..999,b}
24	Viviendas particulares donde no hay residentes habituales	{000..999,b}

25	Total de viviendas particulares	{000..999,b}
26	Viviendas con entrevista completa	{000..999,b}
27	Viviendas con entrevista incompleta	{000..999,b}
28	Viv. con ausencias de residentes al momento de la vista	{000..999,b}
29	Viviendas donde se negaron a dar información	{000..999,b}
30	Viviendas deshabitadas	{000..999,b}
31	Viviendas de uso temporal	{000..999,b}
32	Viviendas donde no hay residentes habituales	{000..999,b}
33	Total de viviendas	{000..999,b}
34	Viviendas colectivas con entrevista completa	{000..999,b}
35	Viviendas colectivas con entrevista incompleta	{000..999,b}
36	Viv. colectivas con ausencias de residentes al momento de la vista	{000..999,b}
37	Viviendas colectivas donde se negaron a dar información	{000..999,b}
38	Total de viviendas colectivas	{000..999,b}

CENSO DE POBLACIÓN Y VIVIENDA 2010
ESTRUCTURA LISTADO DE INMUEBLES
VARIABLES DE INMUEBLES

TR_INMUEBLE

Cons.	Descripción	Rango Válido
1	Identificación de inmueble	{1..99999999}
2	Identificación de listado	{1..9999999}
3	Número de renglón (automático)	{01..20}
4	Código de confronta	{1..3,b}
5	Cotejador	{001..999,b}
6	Consecutivo del inmueble	{001..999,b}
7	Dirección del inmueble (calle, avenida, calzada, privada, carretera...)	{Alfanumérico,b}
8	Número visible (exterior)	{Alfanumérico,b}
9	Número visible (interior)	{Alfanumérico,b}
10	Total de viviendas en el predio	{000..999,b}
11	Total de locales en el predio	{000..999,b}
12	Verificación de residentes	{1,3,b}
13	Tipo de inmueble	{1..8,b}
14	Resultado de la visita (primera)	{1..7,b}
15	Resultado de la visita (segunda)	{1..7,b}
16	Resultado de la visita (tercera)	{1..7,b}
17	Resultado de la visita (supervisor)	{1..7,b}
18	Resultado de la visita (final)	{1..7,b}
19	Registro para localizar a los residentes	{Alfanumérico,b}
20	Código de verificación	{1..7,b}

CENSO DE POBLACIÓN Y VIVIENDA 2010
ESTRUCTURA CUADERNILLO DE POSENUMERACIÓN
VARIABLES DE CARÁTULA

TR_CUADERNILLO_POS

Cons.	Descripción	Rango Válido
1	Identificación de cuadernillo	{1..9999999}
2	Identificación de segmento	{1..9999999}
3	Número de cuadernillo	{00..99}
4	Total de cuadernillos	{00..99}
5	Total de cédulas utilizadas en el cuadernillo	{00..99,b}
6	Folio de paquete	{101..999}
7	Consecutivo del cuadernillo en el paquete	{001..999}
8	Clave del posenumerador(a)	{0..9,b}
9	Clave de supervisor de posenumeradores	{0..9,b}

CENSO DE POBLACIÓN Y VIVIENDA 2010
ESTRUCTURA CUADERNILLO DE POSENUMERACIÓN
VARIABLES DE VIVIENDA

TR_VIVIENDA_POS

Cons.	Descripción	Rango Válido
1	Identificación de vivienda	{1..9999999}
2	Identificación de cuadernillo	{1..9999999}
3	Apellido	{0..9}
4	Consecutivo de la vivienda	{001..999}
5	Número de cédula	{01..99}
6	Total de cédulas en la vivienda	{01..99}
7	Fecha de levantamiento (Día)	{01..31,b}
8	Fecha de levantamiento (Mes)	{05..07,b}
9	Etiqueta	{1..5,b}
10	Consecutivo de la vivienda enumeración	{001..999,b}
11	Total de cuestionarios de la vivienda enumeración	{00..99,b}
12	Tipo de cotejo	{1,2,b}
13	Número del informante	{00..99,b}
14	Verificación de la lista	{1,3,b}
15	Salida de residentes	{1,3,b}
16	Bandera para liberar documentos sin información	{0,1}

CENSO DE POBLACIÓN Y VIVIENDA 2010
ESTRUCTURA CUADERNILLO DE POSENUMERACIÓN
VARIABLES DE PERSONA

TR_PERSONA_POS

Cons.	Descripción	Rango Válido
1	Identificación de persona	{1..99999999}
2	Identificación de vivienda	{1..9999999}
3	Tipo de residente	{1,2}
4	Número automático de persona	{1..6}
5	Código de cotejo	{1..6,b}
6	Número de renglón	{00..99,b}
7	Sexo	{1,3,b}
8	Edad	{000..999,b}
9	Parentesco	{1..9,b}
10	Condición de residencia	{1,3,b}
11	Fecha de llegada (Día)	{00..99,b}
12	Fecha de llegada (Mes)	{00..99,b}
13	Motivo de no residencia	{1..3,b}
14	Fecha de salida (Día)	{00..99,b}
15	Fecha de salida (Mes)	{00..99,b}

CENSO DE POBLACIÓN Y VIVIENDA 2010

ESTRUCTURA PAQUETE

VARIABLES DE PAQUETE

TR_PAQUETE

Cons.	Descripción	Rango Válido
1	Identificación de paquete	{1..9999999}
2	Identificación de segmento	{1..9999999}
3	Folio de paquete	{101..999}
4	Tipo de documento	{1..6}
5	Área operativa (operativo regular o especial)	{000001..999999,b}
6	Área operativa (operativo regular o especial)	{0..9,b}
7	Consecutivo de paquete (operativo de verificación)	{01..99}
8	Total de paquetes (operativo de verificación)	{01..99}
9	Total de listados de inmuebles en el paquete (op. de verificación)	{00..99,b}
10	Total de documentos en el paquete (operativo de verificación)	{000..999,b}
11	Área operativa (operativo de verificación)	{0001..9999,b}
12	Ubicación del paquete en almacén	{0001..9999; A..Z;1..5;1..8}

CENSO DE POBLACIÓN Y VIVIENDA 2010
ESTRUCTURA CUESTIONARIO DEL ENTORNO URBANO
VARIABLES DE CARÁTULA

TR_CUESTIONARIO_URB

Cons.	Descripción	Rango Válido
1	Identificación de cuestionario urbano	{1..9999999}
2	Identificación de segmento	{1..9999999}
3	Apellido	{0..9}
4	Número de cuestionario en la manzana	{01..99}
5	Total de cuestionarios en la manzana	{01..99}
6	Folio de paquete	{101..999}
7	Consecutivo del cuestionario en el paquete	{001..999}
8	Fecha de recorrido (día)	{01..31,b}
9	Fecha de recorrido (mes)	{05..07,b}
10	Clave de supervisor(a)	{0..9,b}
11	Clave de validador(a)	{0..9,b}
12	Manzana de conjunto habitacional	{1,3,b}
13	Acceso a la manzana (toda la manzana)	{1,2,b}
14	Acceso a la manzana (alguna vialidad)	{3,4,b}
15	Acceso a la manzana (edificios o conjuntos habitacionales)	{1,2,b}
16	Precauciones (personas que obstaculizan)	{1,2,b}
17	Precauciones (personas ebrias o drogadas)	{3,4,b}
18	Precauciones (vigilancia)	{1,2,b}
19	Bandera para liberar documentos sin información	{0,1}
20	Observaciones	{Alfanumérico,b}

CENSO DE POBLACIÓN Y VIVIENDA 2010
ESTRUCTURA CUESTIONARIO DEL ENTORNO URBANO
VARIABLES DE VIALIDAD

TR_VIALIDAD_URB

Cons.	Descripción	Rango Válido
1	Identificación de vialidad	{1..9999999}
2	Identificación de cuestionario urbano	{1..9999999}
3	Número automático de la vialidad	{1..6}
4	Nombre de la vialidad	{Alfanumérico,b}
5	Tipo de vialidad o rasgo	{1..4,b}
6	Acceso a personas	{1..3,b}
7	Acceso de automóviles	{1,2,b}
8	Recubrimiento de la calle	{1..3,b}
9	Letrero con nombre de calle (señalización)	{1..4,b}
10	Alumbrado público	{1..4,b}
11	Teléfono público	{1..4,b}
12	Alcantarillas de drenaje pluvial	{1..4,b}
13	Banqueta	{1..4,b}
14	Guarnición	{1..4,b}
15	Árboles o plantas de ornato	{1..4,b}
16	Rampas para silla de ruedas	{1..4,b}
17	Transporte colectivo	{1..4,b}
18	Puestos de comercio semifijo	{1..4,b}
19	Puestos de comercio ambulante	{1..4,b}

CENSO DE POBLACIÓN Y VIVIENDA 2010
ESTRUCTURA CUESTIONARIO DE LOCALIDAD
VARIABLES DE CUESTIONARIO DE LOCALIDAD

TR_CUESTIONARIO_LOC

Cons.	Descripción	Rango Válido
1	Identificación de cuestionario de localidad	{1..9999999}
2	Identificación de segmento	{1..9999999}
3	Folio de paquete	{101..999}
4	Apellido	{0..9}
5	Consecutivo del cuestionario en el paquete	{001..999}
6	Clave de supervisor	{0..9,b}
7	Clave de validador	{0..9,b}
8	Fecha del levantamiento (día)	{01..31,b}
9	Fecha del levantamiento (mes)	{05..07,b}
10	Localidad señalada para llenado	{1,3,b}
11	Situación de la localidad	{1,3,b}
12	Código del inmueble (en ruinas)	{1,b}
13	Total de inmuebles estimados (en ruinas)	{00..99,b}
14	Código del inmueble (en construcción)	{2,b}
15	Total de inmuebles estimados (en construcción)	{00..99,b}
16	Código del inmueble (local con actividad)	{3,b}
17	Total de inmuebles estimados (local con actividad)	{00..99,b}
18	Código del inmueble (local vacío)	{4,b}
19	Total de inmuebles estimados (local vacío)	{00..99,b}
20	Código del inmueble (vivienda)	{5,b}
21	Total de inmuebles estimados (vivienda)	{00..99,b}
22	Total de inmuebles estimados (otra)	{00..99,b}
23	Tipo de inmuebles (descripción otra)	{Alfanumérico,b}
24	Nombre	{Alfanumérico,b}

25	Sexo	{1,3,b}
26	Edad	{00..99,b}
27	Puesto	{1..5,b}
28	Puesto (Otro)	{Alfanumérico,b}
29	Tiempo de residencia (años)	{00..99,b}
30	Tiempo de residencia (No vive en esta localidad)	{00,b}
31	Escolaridad (grado)	{0..9,b}
32	Escolaridad (nivel académico)	{0..4,b}
33	Bandera para liberar documentos sin información	{0,1}
34	Carretera	{1..3,b}
35	Distancia a carretera	{000..999,b}
36	Tiempo de traslado (Horas)	{00..99,b}
37	Tiempo de traslado (Minutos)	{00..99,b}
38	Tiempo de traslado (cabecera municipal o no hay transporte)	{88,89,b}
39	Transporte público	{1..6,b}
40	Frecuencia	{00..99,b}
41	Costo	{000..999,b}
42	Red de agua potable	{1,3,b}
43	Cobro de agua	{1,3,b}
44	Red de drenaje	{1,3,b}
45	Destino del drenaje	{1..4,b}
46	Limpieza y recolección de basura (limpieza de calles)	{1,2,b}
47	Limpieza y recolección de basura (recolección de basura de las viviendas)	{3,4,b}
48	Destino de la basura	{1..5,b}
49	Equipamiento (alumbrado público)	{1,2,b}
50	Equipamiento (calles pavimentadas)	{3,4,b}
51	Equipamiento (pozo de agua comunitario)	{1,2,b}
52	Equipamiento (plaza o jardín)	{3,4,b}
53	Equipamiento (panteón)	{1,2,b}
54	Servicios públicos (oficina o agencia municipal?)	{1,2,b}
55	Servicios públicos (oficina de registro civil?)	{3,4,b}
56	Servicios públicos (policía preventiva?)	{1,2,b}
57	Servicios públicos (cárcel o calabozo?)	{3,4,b}
58	Telecomunicaciones (teléfono público?)	{1,2,b}
59	Telecomunicaciones (caseta de teléfono de larga distancia?)	{3,4,b}
60	Telecomunicaciones (internet público?)	{1,2,b}

61	Comercio (mercado fijo?)	{1,2,b}
62	Comercio (tianguis?)	{3,4,b}
63	Comercio (tienda de abarrotes?)	{1,2,b}
64	Comercio (farmacia?)	{3,4,b}
65	Comercio (papelería?)	{1,2,b}
66	Comercio (tienda o ropa de calzado?)	{3,4,b}
67	Comercio (tienda de muebles o aparatos electrodomésticos?)	{1,2,b}
68	Comercio (tienda de materiales de construcción?)	{3,4,b}
69	Comercio (venta de gas?)	{1,2,b}
70	Abasto de alimentos (frijol?)	{1,2,b}
71	Abasto de alimentos (maíz o harina de maíz)	{3,4,b}
72	Abasto de alimentos (arroz?)	{1,2,b}
73	Abasto de alimentos (leche?)	{3,4,b}
74	Abasto de alimentos (carne de res o de puerco?)	{1,2,b}
75	Abasto de alimentos (pollo?)	{3,4,b}
76	Abasto de alimentos (pescado fresco?)	{1,2,b}
77	Servicios comunales (oficina para enviar y recibir dinero?)	{1,2,b}
78	Servicios comunales (casa de huéspedes?)	{3,4,b}
79	Servicios comunales (hotel o motel?)	{1,2,b}
80	Servicios comunales (fonda?)	{3,4,b}
81	Servicios comunales (restaurante?)	{1,2,b}
82	Servicios comunales (cantina, bar o depósito de licores?)	{3,4,b}
83	Salud (Clínica o centro de salud?)	{1,2,b}
84	Salud (consultorio médico particular?)	{3,4,b}
85	Salud (partera o comadrona?)	{1,2,b}
86	Salud (curandera(o)?)	{3,4,b}
87	Escuelas (preescolar?)	{1,2,b}
88	Escuelas (primaria completa?)	{3,4,b}
89	Escuelas (primaria incompleta?)	{1,2,b}
90	Escuelas (telesecundaria o secundaria?)	{3,4,b}
91	Escuelas (preparatoria o bachillerato?)	{1,2,b}
92	Recintos culturales y deportivos (biblioteca?)	{1,2,b}
93	Recintos culturales y deportivos (casa de la cultura?)	{3,4,b}
94	Recintos culturales y deportivos (cancha deportiva?)	{1,2,b}
95	Lengua indígena	{1,3,b}
96	Nombre de la lengua indígena	{Alfanumérico,b}
97	Lugar de uso (la escuela?)	{1,2,b}
98	Lugar de uso (la iglesia?)	{3,4,b}
99	Lugar de uso (las festividades?)	{1,2,b}
100	Lugar de uso (el hogar?)	{3,4,b}

101	Lugar de uso (para comprar o vender?)	{1,2,b}
102	Actividades económicas 1	{Alfanumérico,b}
103	Actividades económicas 2	{Alfanumérico,b}
104	Actividades económicas 3	{Alfanumérico,b}
105	Conflictos (la propiedad de la tierra?)	{1,2,b}
106	Conflictos (preferencias religiosas?)	{3,4,b}
107	Conflictos (preferencias electorales?)	{1,2,b}
108	Conflictos (delincuencia?)	{3,4,b}
109	Conflictos (alcoholismo o drogadicción?)	{1,2,b}
110	Problema principal	{Alfanumérico,b}
111	Observaciones	{Alfanumérico,b}

Identificación de áreas de oportunidad en tratamiento y tecnologías de información.

Los controles que se llevan actualmente para garantizar la calidad en el proceso ETL no son suficientes para tal efecto, los volúmenes de información que se procesan requieren la implementación de un modelo de calidad de software que permita dar una pauta clara y eficiente pues la inversión de recursos humanos, tecnológicos, y tiempo eleva los costos al implementar estrategias poco planeadas para solventar las situaciones que se han presentado en las diferentes pruebas integrales del procesamiento. La inserción de un modelo de calidad de software como parte del tratamiento de la información representaría un proceso más en el flujo que se lleva a cabo, sin embargo garantizaría en cada etapa la Extracción, Transformación y Carga de la información de las coordinaciones municipales a oficinas centrales y posteriormente al esquema para el procesamiento de la codificación, validación, actualización cartográfica y explotación de la información.

4.3 Modelos de calidad de software.

Algunos investigadores del área de Ingeniería de Software, han modelado el proceso de desarrollo de software utilizando técnicas orientadas a objetos [13]. Otra alternativa es utilizar estándares y prácticas aceptadas por la comunidad como base para introducir el proceso de desarrollo de software [14].

A este conjunto de prácticas responden los modelos de madurez de capacidades que deben ser implantados por cualquier entidad que desarrolle o mantenga software. Estos modelos tienen el propósito de proporcionar recomendaciones a la comunidad de software acerca de lo que debe hacerse para lograr proyectos exitosos y mantener la mejora continua en los procesos [15].

Los modelos de calidad que la industria de la Ingeniería de software ha apoyado para su desarrollo y uso y que como parte de sus recomendaciones incluyen las relacionadas con el proceso ETL son:

Integración de Modelos de Madurez de Calidad (CMMI).

Desarrollados por la Universidad de Carnegie-Mellon han servido de guía para identificar y priorizar las acciones que se realizan para la interpretación de los procesos de ingeniería. Toma en cuenta las prácticas que han sido probadas dentro de la estructura de la organización y ayuda a lograr una madurez y capacidad en las áreas de proceso.

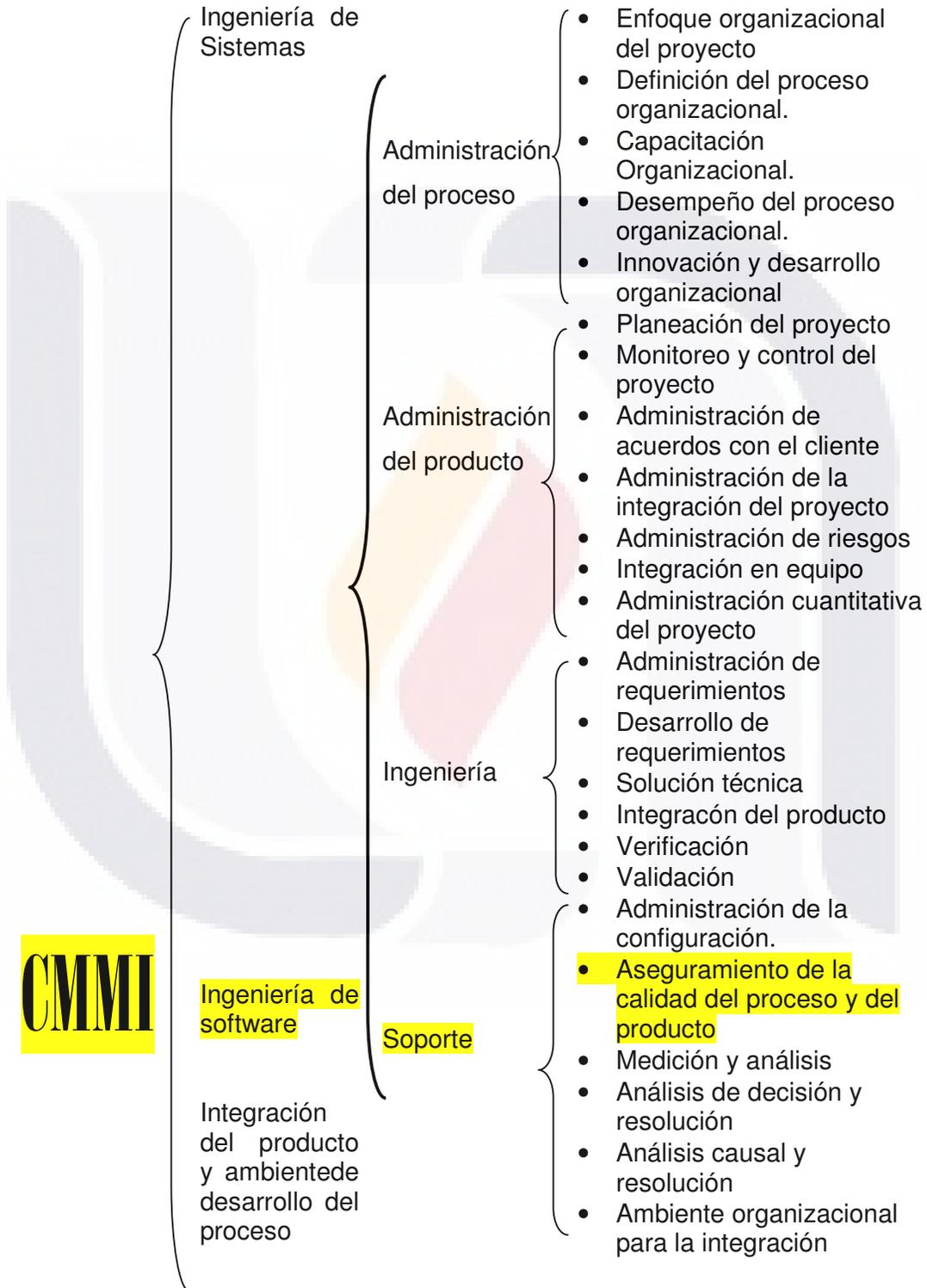
En sus etapas se refiere al desarrollo de requerimientos de los cuales se identifican 3 tipos: los requerimientos del cliente, los requerimientos orientados al producto y los requerimientos orientados a los componentes del producto, en todos ellos es de relevancia la participación de los stakeholders.

Los requerimientos son la base del diseño. El desarrollo de los requerimientos incluye las siguientes actividades:

- Licitación, validación, comunicación, para comprender las expectativas del cliente, así como las restricciones y satisfacer a todos los stakeholders.
- Integrar y coordinar a los stakeholders necesarios.
- Desarrollo del ciclo de vida de los requerimientos del producto.
- Establecimiento de los requerimientos del cliente.

- Establecimiento de los requerimientos del producto inicial y de los componentes del producto alineados a los requerimientos del cliente.

El marco en el cual se ubica el proceso ETL dentro de este modelo se muestra en la siguiente fig. 5 Vista global del modelo CMMI [11]:



Moprosoft.

Las mejores prácticas o modelos de calidad del software han sido diseñados para grandes empresas, por tanto en México esto es un problema, ya que casi el 80% de su industria de software está compuesta por PyMEs, así pues la aplicación de dichas prácticas o modelos ha sido mínima. Moprosoft está fundamentado en el modelo SW-CMM, el estándar ISO 9000 y el reporte técnico ISO/IEC TR 15504, representa el acceso a las prácticas de Ingeniería de Software de clase mundial [17].

Con respecto al proceso de desarrollo y mantenimiento de software clasificado como proceso OPE.2 dentro de la categoría: Operación (OPE), Moprosoft menciona el propósito como la realización sistemática de las actividades de análisis, diseño, construcción, integración y pruebas de productos de software nuevos o modificados cumpliendo con los requerimientos especificados. Para esto sugiere la realización de varios ciclos de desarrollo con las siguientes fases cada uno de ellos:

- Inicio
- Requerimientos
- Análisis y diseño
- Construcción
- Integración y pruebas
- Cierre

Además en cada fase se generan productos que requieren la realización de 3 actividades básicas a saber:

- Distribución de tareas.
- Producción, verificación, validación o prueba de los productos, así como su corrección correspondiente,
- Generación del reporte de actividades.

Entre otros, uno de los objetivos que se establecen en esta parte del proyecto se refieren a:

Lograr que los productos de salida sean consistentes con los productos de entrada en cada fase de un ciclo de desarrollo mediante las actividades de verificación, validación o prueba.

El indicador para este objetivo propone efectuar en cada fase de un ciclo todas las actividades de verificación, validación o prueba, así como las correcciones correspondientes.

Además se definen las siguientes entradas:

Nombre	Fuente
Plan de desarrollo <ul style="list-style-type: none"> • Descripción del producto • Entregables • Proceso Específico • Equipo de trabajo • Calendario 	Administración de proyectos específicos

Y las siguientes salidas:

Nombre	Descripción	Destino
Especificación de requerimientos	Se compone de una introducción y una descripción de requerimientos. Introducción Descripción general del software y su uso en el ámbito de negocio del cliente. Descripción de requerimientos: <ul style="list-style-type: none"> • Funcionales: Necesidades establecidas que debe satisfacer el software cuando es usado en condiciones específicas. Las funcionalidades deben ser adecuadas, exactas y seguras.	Administración de proyectos específicos

	<ul style="list-style-type: none"> • Interfaz con usuario: Definición de aquellas características de la interfaz de usuario que permiten que el software sea fácil de entender, aprender, que genere satisfacción y con el cual el usuario pueda desempeñar su tarea eficientemente. Incluyendo la descripción del prototipo de la interfaz. • Interfaces externas: Definición de las interfaces con otro software o con hardware. • Confiabilidad: Especificación del nivel de desempeño del software con respecto a la madurez, tolerancia a fallas y recuperación. • Eficiencia: Especificación del nivel de desempeño del software con respecto al tiempo y a la utilización de recursos. • Mantenimiento: Descripción de los elementos que facilitarán la comprensión y realización de las modificaciones futuras del software. • Portabilidad: Descripción de las características del software que permitan su transferencia de un ambiente a otro. • Restricciones de diseño y construcción: Necesidades impuestas por el cliente. • Legales y reglamentarios: Necesidades impuestas por leyes, reglamentos, entre otros 	
--	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--

Análisis y diseño	<p>Este documento contiene la descripción textual y gráfica de la estructura de los componentes de software. El cual consta de las siguientes partes:</p> <p>Arquitectónica: Contiene la estructura interna del sistema, es decir la descomposición del sistema en subsistemas. Así como la identificación de los componentes que integran los subsistemas y las relaciones de interacción entre ellos.</p> <p>Detallada: Contiene el detalle de los componentes que permita de manera evidente la construcción y prueba en el ambiente de programación</p>	Administración de proyectos específicos.
Componente	Conjunto de unidades de código relacionadas.	Administración de proyectos específicos.
Software	Sistema de software, destinado a un cliente o usuario, constituido por componentes agrupados en subsistemas, posiblemente anidados.	Administración de proyectos específicos.
Configuración de software	<p>Conjunto consistente de productos de software, que incluye:</p> <ul style="list-style-type: none"> • Especificación de requerimientos. • Análisis y Diseño • Software • Registro de rastreo • Plan de pruebas de sistema 	Administración de proyectos específicos.

	<ul style="list-style-type: none"> • Reporte de pruebas de sistema • Plan de pruebas de integración • Reporte de pruebas de integración • Manual de usuario • Manual de operación • Manual de mantenimiento 	
Manual de usuario	Documento electrónico o impreso que describe la forma de uso del software con base a la interfaz del usuario. Éste deberá ser redactado en términos comprensibles a los usuarios.	Administración de proyectos específicos.
Manual de operación	Documento electrónico o impreso que contenga la información indispensable para la instalación y administración del software, así como el ambiente de operación (sistema operativo, base de datos, servidores, etc) Éste deberá ser redactado en términos comprensibles al personal responsable de la operación	Administración de proyectos específicos.
Manual de mantenimiento	Documento electrónico o impreso que describe la Configuración del Software y el ambiente usado para el desarrollo y pruebas (compiladores, herramientas de análisis y deiseño, construcción y pruebas). Éste deberá ser redactado en términos comprensibles al personal de mantenimiento.	Administración de proyectos específicos.
Reporte de actividades	Registro periódico de actividades, fechas de inicio y fin, responsables y mediciones, tales como:	Administración de proyectos específicos.

	<ul style="list-style-type: none"> • Tiempo de producción, de corrección, de verificación y de validación, • defectos encontrados en verificación, validación o prueba, • tamaño de productos 	
Lecciones aprendidas	Registro de mejores prácticas, problemas recurrentes y experiencias exitosas en la solución de problemas encontrados en un ciclo de desarrollo y mantenimiento.	Conocimiento de la organización
Reporte de mediciones y sugerencias de mejora	<p>Registro que contiene:</p> <ul style="list-style-type: none"> • Mediciones de los indicadores del proceso de Desarrollo y Mantenimiento de software • Sugerencias de mejora al proceso de desarrollo y mantenimiento de software (métodos, herramientas, formatos, estándares, etc.). 	Administración de proyectos específicos.
Registro de rastreo	Relación entre los requerimientos, elementos análisis y diseño, componentes y planes de pruebas.	Administración de proyectos específicos.
Plan de pruebas de Sistema	Identificación de pruebas requeridas para el cumplimiento de los requerimientos especificados.	Administración de proyectos específicos.
Reporte de pruebas de sistema	Registro de participantes, fecha, lugar, duración y defectos encontrados	Administración de proyectos específicos.
Plan de pruebas de integración	<p>Descripción que contiene:</p> <ul style="list-style-type: none"> • El orden de integración de los componentes o subsistemas, guiado 	Administración de proyectos específicos.

	<p>por la parte electrónica del Análisis y Diseño.</p> <ul style="list-style-type: none"> • Pruebas que se aplicarán para verificar la interacción de los componentes. 	
Reporte de pruebas de integración	Registro de participantes, fecha, lugar, duración y defectos encontrados	Administración de proyectos específicos.
Reporte(s) de Verificación	Registro de participantes, fecha, lugar, duración y defectos encontrados	Administración de proyectos específicos.
Reporte(s) de Validación	Registro de participantes, fecha, lugar, duración y defectos encontrados	Administración de proyectos específicos.

SPICE (Software Process Improvement and Capability dEtermination).

Las prácticas que define sirven como una actividad esencial de un proceso específico y se agrupan por área y propósito de la actividad. Para este estudio el propósito de la actividad se centra en garantizar la calidad en la información al aplicar el proceso ETL sobre información censal.

El modelo propone varios aspectos interesantes en el desarrollo de proyectos de software abordando en la categoría de Ingeniería los siguientes procesos:

ENG. 1 Diseño y desarrollo de requerimientos del sistema

ENG. 2 Desarrollo de los requerimientos de software

ENG. 3 Desarrollo del diseño de software

ENG. 4 Implementación del diseño de software

ENG. 5 Integración y pruebas del software

ENG. 6 Integración y pruebas del sistema

ENG. 7 Mantenimiento del sistema y del software

TESIS TESIS TESIS TESIS TESIS

Inmerso en estos procesos se encuentra nuestro objeto de estudio, los factores relacionados con el proceso de Extracción, transformación y carga de la información como parte de los aspectos que se consideran en el diseño y desarrollo del software.

Posteriormente el modelo SPICE se refiere a la gestión de la calidad como proceso para asegurar que los productos o servicios satisfagan al cliente o usuario. Para ello se consideran los siguientes procesos:

- PRO.5.1 Establecer metas de calidad.
- PRO.5.2 Definir las métricas de calidad.
- PRO.5.3 Identificar actividades de calidad.
- PRO.5.4 Realizar las actividades de calidad.
- PRO.5.6 Tomar las acciones correctivas

V. PROPUESTA DE MODELO DE CALIDAD DE SOFTWARE PARA EL PROCESO ETL DE INFORMACIÓN CENSAL.

La propuesta del modelo de calidad presentada en este trabajo se trata de una adaptación basada en los modelos que han sido utilizados en grandes y medianas empresas, así como la experiencia propia en el INEGI quien tiene la facultad exclusiva de realizar los censos de población y vivienda en México y quien en primera instancia sería el usuario del modelo de calidad.

En el modelo se recogen todas las actividades relacionadas con la calidad y específicamente al proceso de extracción, transformación y carga de información, aunque en el análisis realizado sobre los modelos, en éstos no se aborda el proceso a ese nivel, sino mas bien sobre el proyecto o producto de software, en esta propuesta es posible identificar los factores necesarios para garantizar la calidad en la información una vez que se pasa por el proceso ETL.

En el modelo, se propone un método formado por un conjunto de actividades que, a su vez, se dividen en tareas. Para cada tarea se identifican los productos de entrada y los productos de salida. También se establecen las técnicas y prácticas útiles en el desarrollo de cada tarea así como los participantes en la misma.

El modelo propuesto divide las categorías de procesos en tres grandes grupos:

La categoría de proceso de calidad. Consistente en el establecimiento de los parámetros generales de calidad para el proceso específico ETL una vez que los stakeholders han decidido obtener información de calidad en el tratamiento de su información y estén de acuerdo en continuar en el camino de la mejora.

La categoría de proceso de extracción. Consistente en el proceso de especificar, implementar o mantener un sistema que permita la identificación de la información importante y necesaria captada en las fuentes originales de datos.

La categoría de proceso de transformación. Consistente en el proceso de identificación y aplicación de métodos de conversión o transformación de

atributos de la información que por su naturaleza y conveniencia para el procesamiento se requiere manipular bajo una estructura diferente a la captada en el proceso de extracción.

La categoría de proceso de carga. Consistente en el proceso de adecuación de estructuras de fuentes datos a bases de datos requeridas por los procesos de tratamiento de información posteriores al proceso ETL.

Cada proceso en este modelo se describe a manera de términos de bases prácticas. La utilización en casos prácticos posicionará a la organización al menos en el nivel 3 de CMMI.



Fig. 6 Categorías de proceso en la propuesta el modelo de calidad

5.1 Definición.

A cada práctica se asigna un identificador de la forma CP.PR.PT donde:

CP Categoría de proceso

PR Número de proceso

PT Número de práctica

Por ejemplo “EXT.3.1” denota la base práctica dentro de la categoría de proceso EXT (Extracción), proceso 3 (Envío-Recepción), base práctica 1 “Envío-Recepción-Integración”.

5.2 Bases prácticas

Todas las categorías de procesos, procesos y bases prácticas se presentan con un nombre y una descripción.

Las tres categorías de procesos son:

CAL Calidad

EXT Extracción

TRA Transformación

CAR Carga

La descripción de cada categoría de procesos incluye una lista de nombres de procesos. Se menciona el propósito de cada proceso, los insumos y las salidas y se describe en qué momento inicia el siguiente proceso. Las bases prácticas se numeran para permitir una fácil identificación, seguido de esto viene una descripción de lo que se debe realizar en la práctica proporcionando un ejemplo o explicación detallada.

CALIDAD (CAL)

La categoría de procesos Calidad consiste en el establecimiento de los parámetros generales de calidad para el proceso específico ETL una vez que los stakeholders han decidido obtener información de calidad en el tratamiento de su información y estén de acuerdo en continuar en el camino de la mejora.

Los procesos que pertenecen a la categoría de procesos son:

CAL.1 Establecer metas de calidad.

CAL.2 Definir métricas de calidad.

CAL.3 Identificar actividades de calidad.

CAL.4 Ejecutar actividades de calidad.

CAL.5 Tomar acciones correctivas.

CAL.1 Establecer metas de calidad.

El propósito de Establecer metas de calidad es definir los factores que se tomarán en cuenta para la revisión del proceso en base a los requerimientos de calidad.

CAL.1.1 Identificar áreas de oportunidad. En base a estudios similares o experiencias en proyectos anteriores propios, determinar los puntos débiles y convertirlos en metas alcanzables que además garanticen calidad en los resultados.

CAL.2 Definir métricas de calidad.

El propósito de Definir métricas de calidad es establecer una unidad de medida por cada meta de calidad establecida en el proceso anterior.

CAL.2.1 Realizar diferentes escalas de medición. Elegir la más adecuada para cada meta.

CAL.3 Identificar actividades de calidad.

El propósito de identificar actividades de calidad es asociar a cada meta establecida las actividades necesarias para satisfacer la meta.

CAL.3.1 Seleccionar actividades que permitan la mejora continua. De las actividades que pudieran realizarse para lograr las metas de calidad establecidas, seleccionar además aquellas que prometan mejora permanente como parte del proceso.

CAL.4 Ejecutar actividades de calidad.

El propósito de Ejecutar actividades de calidad es experimentar las actividades identificadas en el proceso anterior sobre el proyecto.

CAL.4.1 Validar el orden de ejecución. Verificar que para ejecutar las actividades de calidad se realice en un orden específico si estas deben llevarlo.

CAL.4.2 Validar que se tengan las condiciones necesarias. Verificar que las personas, las herramientas, la infraestructura, los procesos, las variables internas y externas sean propicios para la ejecución de las actividades de calidad.

CAL.5 Tomar acciones correctivas.

El propósito de tomar acciones correctivas es considerar que algunas veces por alguna razón la meta de calidad no es satisfecha y se vuelve necesario revisar las alternativas disponibles.

CAL5.1 Identificar la razón por la cual la meta de calidad no fue satisfecha. Para tomar la acción correctiva necesaria.

EXTRACCIÓN (EXT)

La categoría de Extracción Consistente en el proceso de especificar, implementar o mantener un sistema que permita la identificación de la información importante y necesaria captada en las fuentes originales de datos.

Los procesos que pertenecen a la categoría de procesos son:

EXT.1 Definir estructuras para almacenamiento de datos.

EXT.2 Identificar instalación de infraestructura.

EXT.3 Envío-recepción.

EXT.1 Definir estructuras para almacenamiento de datos.

El propósito de Definir estructuras para almacenamiento de datos es mantener una estructura para alojar la información importante en tablas de base de datos.

EXT.1.1 Identificación de entidades. Temas importantes que agrupan características propias de la información captada.

EXT.1.2 Esquematizar entidades y relaciones. En un diagrama entidad – relación o modelo de datos para apoyarse y tener en cuenta la relación existente entre cada entidad identificada.

EXT.1.3 Elaborar diccionarios de datos. Que contienen el nombre, la descripción, el tipo y la longitud de cada dato de la entidad identificada, así como los valores válidos,

EXT.1.4 Crear estructuras de datos. En base al modelo y diccionario de datos crear las tablas con sus relaciones.

EXT.2 Identificar instalación de infraestructura.

El propósito de Identificar instalación de infraestructura consiste en asegurar que los equipos y servidores trabajen en sincronía para el correcto flujo y manejo de información.

EXT.2.1 Revisar funcionamiento de equipo de comunicaciones. Consiste en probar flujos de volúmenes de información similares a los que se procesarán.

EXT.2.2 Revisar configuración de servidores. De acuerdo a las necesidades propias de procesamiento se debe hacer una revisión física y lógica que se ajuste a ellas.

EXT.3 Envío-recepción.

El propósito de la categoría de proceso Envío-recepción consiste en asegurar que se cumple con los estándares y normas de comunicación así como la integridad de la información.

EXT.3.1 Envío-recepción-integración. Asegurar la integridad de la información al momento de ser recibida comparada con la enviada (Ver Anexos).

TRANSFORMACIÓN (TRA)

La categoría de proceso de transformación. Consistente en el proceso de identificación y aplicación de métodos de conversión o transformación de atributos de la información que por su naturaleza y conveniencia para el procesamiento se requiere manipular bajo una estructura diferente a la captada en el proceso de extracción.

Los procesos que pertenecen a la categoría de procesos son:

TRA.1 Identificar atributos de entidades que requieren proceso de transformación.

TRA.2 Establecer reglas de transformación por atributo.

TRA.3 Desarrollar software para ejecutar reglas de transformación.

TRA.4 Ejecutar software que ejecuta reglas de transformación.

TRA.5 Evaluación de proceso de transformación.

TRA.1 Identificar atributos de entidades que requieren proceso de transformación.

El propósito de Identificar atributos de entidades que requieren proceso de transformación es justamente considerar solo aquellos atributos de entidades que por su naturaleza no pueden ser solamente extraídos sino pasar por un proceso de transformación para aprovechar la nueva estructura del dato y procesarlo en etapas posteriores.

TRA.1.1 Revisar la manera en que se capta cada atributo de las entidades. Permite conocer la estructura original de cada atributo de cada entidad.

TRA.1.2 Considerar la importancia del atributo con respecto al resto de atributos al interior de la entidad.

TRA.2 Establecer reglas de transformación por atributo.

El propósito de Establecer reglas de transformación por atributo es tener claro el proceso de transformación que se requiere con el fin de optimizar tiempos y recursos para el procesamiento de la información.

TRA.2.1 Analizar la información. Para determinar la conveniencia de aplicar la transformación de la mejor manera.

TRA.2.2 Elaborar catálogos. Necesarios para apoyar el proceso de transformación basado en reglas definidas.

TRA.2.3 Documentar proceso. Indicando estrategias generales (todas las tablas de la base de datos) y particulares (por cada atributo)

TRA.3 Desarrollar software para ejecutar reglas de transformación.

El propósito de Desarrollar software para ejecutar reglas de transformación es automatizar el proceso para generar la información transformada en tiempos óptimos y por lo tanto oportunos para dar seguimiento al flujo de información.

TRA.3.1 Elaborar herramientas de apoyo para el desarrollo (diagramas)

TRA.3.2 Considerar buenas prácticas de programación. Como auto-documentación de código, etc.

TRA.3.3 Seguir una metodología de desarrollo probada para generar documentación de pruebas.

TRA.4 Ejecutar software que ejecuta reglas de transformación.

El proceso de Ejecutar software que ejecuta reglas de transformación tiene el propósito de obtener la información como se requiere en el procesamiento, es decir, eliminando, combinando o transformando atributos de la base de datos.

TRA.4.1 Preparar información de base de datos. En el caso de realizar pruebas para controlar los resultados de la transformación.

TRA.4.2 Preparar rutinas automáticas que permitan el registro de resultados parciales en el proceso automático de transformación.

TRA.5 Evaluación de proceso de transformación.

El proceso de Evaluación de proceso de transformación tiene el propósito de entregar información que permita decidir si el proceso de transformación en general está cumpliendo los requerimientos.

TRA.5.1 Analizar los resultados. Registrados en la categoría de proceso anterior.

TRA.5.2 Tomar acciones correctivas. En caso de observar un desempeño de transformación que no satisface las pruebas controladas o los resultados esperados.

CARGA (CAR)

La categoría de proceso de carga. Consistente en el proceso de adecuación de estructuras de fuentes datos a bases de datos requeridas por los procesos de tratamiento de información posteriores al proceso ETL.

Los procesos que pertenecen a la categoría de procesos son:

CAR.1 Establecer reglas de integridad.

CAR.2 Definir estrategia de bases de datos de “paso”.

CAR.3 Definir estrategias de estados de la información.

CAR.1 Establecer reglas de integridad.

El proceso de Establecer reglas de integridad tiene como propósito identificar las relaciones entre tablas de la base de datos que servirá para el tratamiento de la información.

CAR.1.1 Esquematizar el modelo de datos. Luego que se realiza el proceso de transformación. A partir de este modelo quedarán claras las reglas de integridad y relaciones para el proceso de carga de información.

CAR.1.2 Documentar diccionario de datos. En base al modelo generado indicando la relación y reglas de integridad por cada tabla de información.

CAR.2 Definir estrategia de bases de datos de “paso”.

El proceso Definir estrategia de bases de datos de “paso” tiene como propósito identificar si son necesarias las estructuras intermedias de datos para llegar a la carga de información en los servidores destino.

TESIS TESIS TESIS TESIS TESIS

CAR.2.1 Analizar estrategia. Tomando en cuenta los recursos tecnológicos que se tienen y el impacto en los tiempos y verificaciones intermedias que se tuviera que realizar durante la generación de estas bases de datos de “paso”.

CAR.2.2 Aceptar o rechazar la estrategia. Dependiendo del análisis realizado en el proceso anterior.

CAR.2.3 Documentar análisis y toma de decisión. Justificando la decisión tomada.

CAR.3 Definir estrategias de estados de la información.

El proceso de Definir estrategias de estados de la información tiene como propósito generar el catálogo de estados por los que la información pasa antes de considerarse “cargada” en los servidores destino.

CAR.3.1 Estudiar los momentos que tiene la información para identificar en cuál de ellos es conveniente asignar un identificador o estatus. Ejemplo:

- 0 Cargando archivo
- 1 Archivo cargado correcto
- 2 Integrande
- 3 Archivo integrado correcto
- 4 Archivo integrado correcto con registros duplicados
- 5 Archivo integrado correcto con error en registros
- 6 Tramo de control bloqueado (se establece de forma manual)

VI. CONCLUSIONES.

Al concluir este trabajo se confirma que hoy por hoy, frente al escenario actual el reto de las tecnologías de información consiste en plantear soluciones y soluciones de calidad a las problemáticas que el mundo del software requiere.

Este trabajo de tesis recopila un esfuerzo por el entendimiento e interpretación de los modelos de calidad de software que aplican grandes compañías con fines específicos. Como resultado de este esfuerzo se presenta una propuesta con la finalidad de proveer un modelo de calidad de software para que sirva como guía en el proceso de Extracción, Transformación y Carga de información a organismos que como una de sus metas tienen la de generar información a partir de censos, tales como el INEGI en nuestro país.

6.1 Contribuciones, ventajas y desventajas de la propuesta.

La propuesta de modelo de calidad contribuye a que empresas que dan tratamiento a información censal consideren que el aspecto de calidad en los procesos necesarios de extracción transformación y carga permite eliminar el cuello de botella que regularmente representa este proceso en el flujo de la información para llegar a la etapa de explotación y difusión de resultados del proyecto censal. Al mismo tiempo permite que las organizaciones a nivel internacional cuenten con una guía que mejorará la calidad de la información durante la planeación y desarrollo de software relacionado con el proceso ETL, de esta manera la propuesta de modelo asegura que no solo la información resultante sea de calidad sino todo el proceso que se sigue para su obtención.

6.2 Evaluación de objetivos

Haciendo referencia al **Objetivo Particular** planteado en el capítulo III de este trabajo, donde se señala “Analizar la tendencia sobre los procesos ETL de diferentes modelos de calidad de software utilizados por pequeñas, medianas y grandes compañías”.

Se obtienen los siguientes resultados:

Resultado: El objetivo se cumple.

Justificación: Se realizó un estudio sobre tres modelos que tienen el propósito de proporcionar recomendaciones a la comunidad de software acerca de lo que debe hacerse para lograr proyectos exitosos y mantener la mejora continua en los procesos (*Ver desarrollo del estudio en capítulo IV*).

El resultado anterior ayuda a responder **afirmativamente** a la pregunta de investigación No. 1 “¿Los modelos de calidad de software utilizados en pequeñas, medianas y grandes compañías orientan parte de sus procesos específicamente a la atención de los procesos ETL?”, debido a que la teoría de los modelos CMMI, SPICE y Moprosoft describen en sus procesos las actividades necesarias para obtener productos y proyectos de software de calidad, poniendo atención en los procesos de validación a los que se somete la información, por lo que se concibe la proposición 1 como **verdadera**

Haciendo referencia al **Objetivo Particular** planteado en el capítulo III de este trabajo, donde se señala “Analizar la metodología que ha utilizado el INEGI para procesar la ETL en los proyectos censales y determinar si la información generada es de calidad para las etapas posteriores del tratamiento”.

Se obtienen los siguientes resultados:

Resultado: El objetivo se cumple.

Justificación: Se analizó la manera en que el INEGI ha llevado a cabo una metodología que le permite realizar el proceso ETL y continuar el tratamiento a su información, sin embargo al interior del instituto no se han realizado estudios

que midan la calidad de la información emitida por el proceso ETL (*Ver desarrollo del estudio en capítulo IV*).

El resultado anterior ayuda a responder **afirmativamente** a la pregunta de investigación No. 2 “¿La metodología utilizada por el INEGI en los eventos censales para el procesamiento ETL ha proporcionado información de calidad para las posteriores etapas del tratamiento?”, debido a que la teoría del estudio realizado sustenta que la información resultante del proceso ETL ha satisfecho las necesidades de tratamiento de información aunque no se hayan realizado estudios específicos para medir la calidad en la información procesada, por lo que se concibe la proposición 2 como **verdadera**.

Haciendo referencia al **Objetivo Particular** planteado en el capítulo III de este trabajo, donde se señala “Proponer, como resultado del análisis, un modelo de calidad de software para el proceso de ETL de información censal”.

Se obtienen los siguientes resultados:

Resultado: El objetivo se cumple.

Justificación: Se elaboró una propuesta de modelo calidad de software orientada al proceso ETL y basada en los modelos CMMI, SPICE, Moprosoft y la metodología utilizada en el INEGI (*Ver propuesta del modelo en el capítulo V*).

El resultado anterior ayuda a responder **afirmativamente** a la pregunta de investigación No. 3 “¿Es factible elaborar una propuesta de modelo de calidad para procesos ETL sobre información censal a partir de estudios de modelos de calidad existentes y/o la metodología utilizada por el INEGI?”, debido a que las teorías base se comprendieron e interpretaron permitiendo la creación de la propuesta creativa, de esta manera se concibe la proposición 3 como **verdadera**.

6.3 Trabajos futuros

Evaluación del modelo de calidad de software en los procesos ETL de información censal.

Para investigadores en el campo de la ingeniería de software resultará atractivo poner en práctica el trabajo propuesto y utilizarlo como insumo para proyectos censales en México y otros países. En México se realizan censos de Población y Vivienda, Económicos y Agropecuarios, además de Conteos de Población con diferente periodicidad pero estructura y estrategias de procesamiento de información similares. De esta manera además de contribuir en la mejora del proceso ETL podría obtener una evaluación del modelo de calidad propuesto.

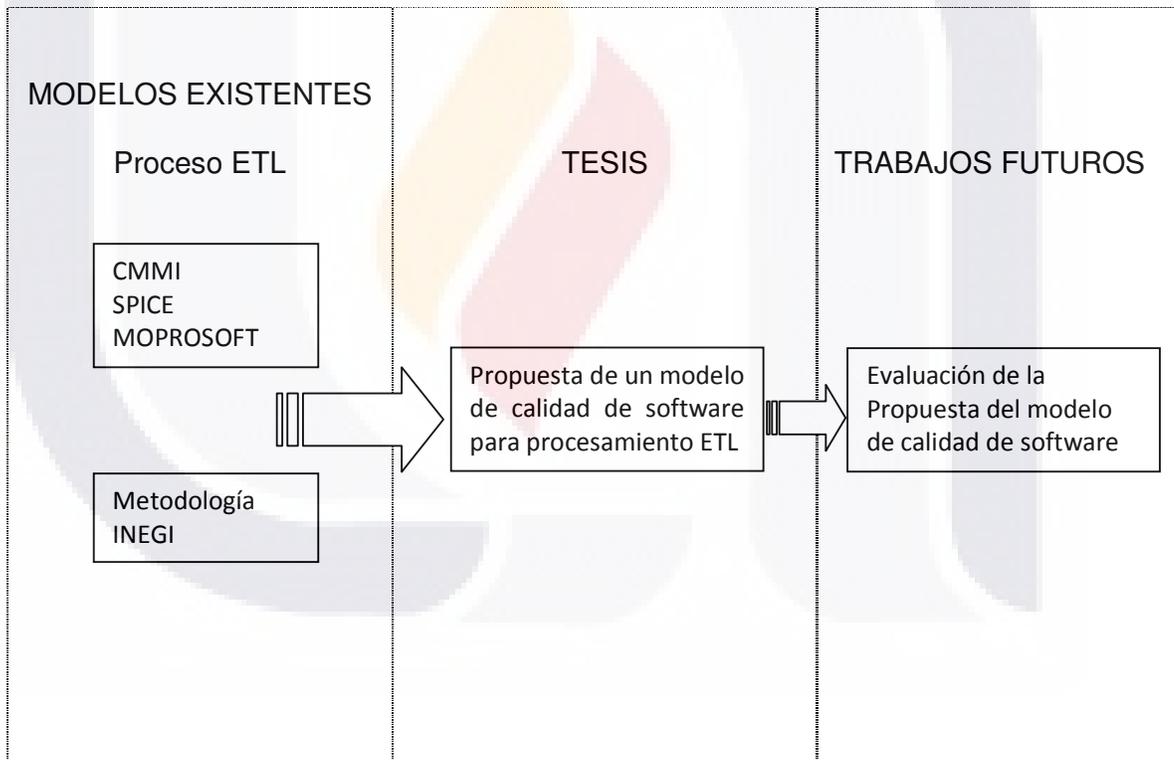


Fig. 7. Esquema para trabajos futuros

VII. MARCO DE REFERENCIA

7.1 Glosario

Extracción. La extracción convierte los datos a un formato preparado para iniciar el proceso de transformación.

Transformación. Aplica una serie de reglas de negocio o funciones sobre los datos extraídos para convertirlos en datos que serán cargados.

Carga. Es el momento en el cual los datos de la transformación son pasados al sistema de destino

Acumulación simple. Es una forma de carga de datos; consiste en realizar un resumen de todas las transacciones comprendidas en el período de tiempo seleccionado y transportar el resultado como una única transacción hacia el destino de datos.

Rolling. Forma de carga que se aplica en los casos en que se opta por mantener varios niveles de granularidad. Información resumida a distintos niveles correspondientes a distintas agrupaciones de la unidad de tiempo o diferentes niveles jerárquicos en alguna o varias de las dimensiones de la magnitud almacenada (por ejemplo, totales diarios, totales semanales, totales mensuales, etc.).

RTF. Revisiones técnicas formales, permiten la revisión del cumplimiento del proceso de software definido, descubrir errores en la función, la lógica o implementación, verificar que el software alcanza sus requisitos.

VIII. ANEXOS

Propuesta de control de envíos, recepción e integración

Administrador Web

Censo de Población y Vivienda 2010.

Para asegurar la consistencia entre la información enviada por el Administrador Local y el Administrador Web es conveniente establecer controles durante el proceso. Los controles que se están considerando son el Acuse de recepción – integración y la verificación de secuencia. Estos controles permiten que todos los archivos de información que se genere en el Administrador Local sean enviados a oficinas centrales y que el orden de recepción sea el correcto para poder aplicar adecuadamente los procesos de inserción y/o actualización.

Proceso de envío, recepción e integración de archivos

El proceso de envío, recepción e integración de archivos servirá de puente de comunicación entre el Administrador Local y el Administrador Web, permitiendo a las Coordinaciones Municipales saber a través de archivo de texto llamado acuse, si el envío realizado se integró a la base de datos nacional o si fue rechazado.

Este proceso comienza cuando la Coordinación Municipal realice su primer envío de información utilizando el Administrador Local, una vez realizado el primer envío, el administrador Web lo recibirá y le aplicará un proceso para determinar si la información del paquete es correcta ó no, en caso de que la información sea correcta, se integrará de manera automática a la base de datos nacional, por otra parte, si la información del paquete es incorrecta, es decir no cumple con ciertos criterios preestablecidos, se rechaza el envío y no se integra la información del paquete a la base de datos nacional.

Independientemente de que la información del archivo sea correcta o no, el Administrador Web deberá generar un Acuse que indique la situación del archivo, si el archivo fue rechazado, el acuse indicará los errores por los cuales fue rechazado, en caso de que el archivo se haya integrado, el acuse mostrará

para cada archivo, la cantidad de registros que fueron integrados a la base de datos nacional.

El Administrador Local ubicado en la Coordinación Municipal, tiene la funcionalidad de recuperar del Administrador Web al disco duro de la Laptop el Acuse que le indica la situación de su envío, en caso de tener el acuse previo, el Administrador Local no permitirá realizar el siguiente envío.

El Administrador Local sólo podrá realizar un envío x (donde $x = 2,3,4,\dots,n$) de información al Administrador Web, siempre y cuando exista en la Laptop el acuse y éste indique que el archivo es correcto ó cuando el acuse indique que es incorrecto y la información del archivo ya este corregida.

Acuse de recepción e integración correcta

Cuando la integración del archivo ZIP ha sido correcta se genera un archivo con la siguiente estructura:

{nombre del archivo de acuse}

Usuario: {nombre del usuario que hizo el envío}

Fecha: {Fecha del envío}

Archivo: {Nombre del archivo de información ZIP}

Tamaño:{Tamaño del archivo}

Envío: {número de envío}

Estatus: {código de estatus}

Resultado del Contenido:

Nombre de la Tabla: {nombre de la tabla 1}

Total Inserciones: {total de inserciones en la tabla 1}

Total Actualizaciones o Duplicados: {total de registros duplicados o actualizados de la tabla 1}

Total Registros Erroneos : {total de registros con error de la tabla 1}

Nombre de la Tabla: {nombre de la tabla 2}

Total Inserciones: {total de inserciones en la tabla 2}

Total Actualizaciones o Duplicados: {total de registros duplicados o actualizados de la tabla 2}

Total Registros Erróneos: {total de registros con error de la tabla
2}

Nombre de la Tabla: {nombre de la tabla n}

Total Inserciones: {total de inserciones en la tabla n}

Total Actualizaciones o Duplicados: {total de registros duplicados
o actualizados de la tabla n}

Total Registros Erroneos : {total de registros con error de la tabla
n}

Nota: En la lista de tablas integradas solo se consideran las tablas de liberado
de procesos primarios.

Acuse de recepción incorrecta

Cuando la integración del archivo ZIP no cumplió con los criterios de recepción
se genera un archivo con la siguiente estructura:

{nombre del archivo de acuse}

Usuario: {nombre del usuario que hizo el envío}

Fecha: {Fecha del envío}

Archivo: { Nombre del archivo de información ZIP }

Tamaño:{Tamaño del archivo}

Envío: {número de envío}

Estatus: {código de estatus}

Lista de errores:

{Descripción del error 1}

{Descripción del error 2}

{Descripción del error n}

Los acuses se generan encriptados con extensión DAT.

Verificación de secuencia

Cada archivo ZIP que se genere en el Administrador Local contendría un
archivo de verificación de secuencia, el cual se llamará archivos.dat (el original
en txt y cifrado de acuerdo a la función).

El archivo de validación tiene la siguiente estructura:

{Archivo anterior generado (si es el primero debe traer: CP.ZIP)} tabulador

{Archivo actual enviado}

{Archivo BD SQLite} tabulador {Fecha última modificación Archivo BD SQLite }

{Archivo BD SQLite respaldo completo} tabulador {Fecha última modificación

Archivo BD SQLite respaldo completo}

La fecha de la última modificación con formato ddMMyyyyHHmms.

Validación del archivo

Para que un envío sea aceptado por el sitio debe cumplir los siguientes criterios:

- Nombre de archivo correcto
- El tramo de control este registrado en la BD central
- Ese mismo archivo no exista ya registrado como “cargado correcto”
- Contenido correcto del archivo ZIP
- El estatus del último archivo recibido e integrado para ese tramo es correcto.

Nombre de tabla correcto

El nombre del archivo ZIP debe tener la estructura que se establezca, por ejemplo:

EECENZM_ENV_DDMMAAAA_HHMM.ZIP

Donde:

EE = Entidad (2 posiciones)

CE = Coordinación Estatal (2 posiciones)

N = Coordinación Enumeradores (1 posición)

Z = Coordinación de Zona (1 posición)

M = Coordinación Municipal (1 posición)

ENV = es un consecutivo de envíos (3 posiciones).

DDMMAAAA_HHMM = es la fecha de generación del archivo ZIP

El archivo de acuse se llamará igual que el del ZIP al cual corresponde, pero con extensión DAT: EECECZM_ENV_DDMMAAAA_HHMM.DAT

El tramo de control este registrado en la BD central

El tramo de control especificado en el archivo debe estar registrado en el catálogo de tramos de control de oficinas centrales.

Ej. Para 0101111_001_15102007_0800.ZIP, debe existir registrado el tramo de control Entidad 01, CE 01, N 1, Z 1 y M 1 en la BD de oficinas centrales.

El envío no exista ya registrado como correcto

No debe existir un envío registrado con el mismo tramo de control y número de envío, de lo contrario, se rechazara el envío, porque ya está registrado

Ej. Si se intenta enviar el archivo 0101111_001_15102007_0800.ZIP pero ya se cuenta con un envío “cargado correcto” registrado con un nombre de tabla 0101111_001_14102007_1355.ZIP el nuevo envío será rechazado.

Contenido correcto del archivo ZIP

Para que el contenido del archivo ZIP de envío sea considerado válido debe cumplir lo siguiente:

- Todas las tablas del modelo de integración deben venir, aunque no contengan registros.
- Los registros recibidos en cada tabla deben coincidir con los registrados en la tabla de TR_TOTALES.
- El nombre actual del ZIP de envío debe coincidir con lo informado en archivos.dat
- El nombre del archivo anterior informado en archivos.dat debe existir registrado como envío “cargado correcto” en la BD central, para el caso del primer archivo enviado, el archivo anterior especificado en archivos.dat debe ser CP.ZIP.

En caso de no cumplir con cualquiera punto de los anteriores, el envío será rechazado.

Notas

El archivo ZIP de información tiene contraseña.

Bibliografía.

1. M.S. Krishnan, Tridas Mukhopadhyay, Dave Zubrow (2000). *Software Process Models and Project Performance*. Boston
2. Buvaneswari K. Venkataraman, William A. Ward, Jr. (June 1999), *An Introduction to Software Quality*
3. *Software Engineering: A Practitioner's Approach, Fifth Edition, 2000*. Roger S. Pressman
4. Pandian C. Ravindranath Pandian (2004). *Software Metrics , A Guide to Planning , Analysis , and Application . Florida, USA. Auerbach*
5. Cernuda del Río, J.E. Labra Gayo, y J.M. Cueva Lovelle, "Verificación y validación mediante un modelo de componentes,"
6. Simitsis, A. & Vassiliadis, P. (2004). *A Methodology for the Conceptual Modeling of ETL Processes*.
7. *Conferencia Estadística de las Américas. CEPAL, Julio 2007. Principios y Recomendaciones para los Censos de Población y Vivienda/Habitación de las Naciones Unidas*.
8. *INFORME DEL SEMINARIO SOBRE ORGANIZACIÓN GERENCIAL Y ESTRATEGIA DE UN CENSO (22/7/2008)*. Lima, Perú, 24 al 28 de octubre de 1988.
9. *Liga de conceptos relacionados con el CPV emitidos en la página del INEGI*.
<http://www.inegi.org.mx/est/contenidos/proyectos/Preview.aspx>
10. *Ib Revista de Información Básica. Revista Virtual ISSN1909-2466 Vol.1 No. 2*.
http://www.dane.gov.co/revista_ib/html_r2/articulo12_r2.htm
11. Beatriz Durón Rosales, María de Lourdes Y. Margain Fuentes. Junio 2004. "Diseño y evaluación de un modelo conceptual para la interpretación de evaluaciones de los procesos de ingeniería del modelo integrado de madurez de capacidades".
12. Ángel Cifuentes Garres, Olalla Cruz Mata, Ana María Lledó Álvarez 2008. Ponencia: "Aseguramiento de la calidad de los datos en el proceso ETL del MNP".

13. *García Mireles, G. y Rodríguez Jacobo, J. (2001). Aplicación del modelado de procesos en un curso de ingeniería de software. Revista Electrónica de Investigación Educativa, 3 (2). Consultado el día de mes de año en: <http://redie.uabc.mx/vol3no2/contenido-mireles.html>*
14. *Maria Letizia Jaccheri, Gian Pietro Picco, Patricia Lago (1998) ACM Transactions on Software Engineering and Methodology (TOSEM) achive, Volume 7 , Issue 4 (October 1998) table of contents. Pages: 368 - 410, ISSN:1049-331X*
15. *Software Engineering Institute (1997).*
16. *D. Enrique Morán Aláez, Dña. Pilar Martínez Rollón, Dña. Pilar Vázquez Sancho (2006). Nuevos métodos en la elaboración de estadísticas censales de población: EPV06*
17. *Hanna Oktaba et al. (Agosto 2005). Modelo de procesos para la industria del Software (MOPROSOFT) ver. 1.3*
18. *Guía de Autoevaluación de la categoría de procesos de Ingeniería de Software bajo el estándar ISO-15504 [SPICE]. M en C Víctor González Castro*
19. *La mejora de los procesos de software en las pequeñas y medianas empresas (pyme). Un nuevo modelo y su aplicación a un caso real (2005) Revista Española de Innovación, Calidad e Ingeniería del Software, Vol.1, No. 2.*
20. *Ph. D.Thesis, Universitat de les Illes Balears (2005). Mas, A. Un Nuevo Modelo para la Implantación de un Sistema de Gestión de Calidad en Pymes de Desarrollo de Software basado en SPICE (ISO/IEC 15504).*