

Departamento Administrativo Nacional de Estadística



**Dirección de Metodología y Producción
Estadística-DIMPE**

**Procesamiento de Datos
Encuesta de
Microestablecimientos
- MICRO**

Abril 2009

	<p align="center">PROCESAMIENTO DE DATOS MICROESTABLECIMIENTOS - MICRO</p>	<p>CÓDIGO: SI-MICRO-MOT-02 VERSIÓN: 2 PAGINA: 2 FECHA: 30-04-09</p>
ELABORÓ: APOYO DE SISTEMAS	REVISÓ: EQUIPO DE SISTEMAS	APROBÓ : DIRECTOR DIMPE

CONTENIDO

1. SISTEMA DE PROCESAMIENTO DE DATOS	2
2. OBJETIVOS	2
2.1 Objetivo General	2
2.2 Objetivos Específicos	2
3. SERVICIOS DE INFORMACIÓN	2
4. MODELO FUNCIONAL DEL SISTEMA	2
4.1 Subproceso de Control de Recepción de Datos.....	2
4.2 Subproceso de Cargue y Validación de Datos	2
4.3 Subproceso de Edición e Imputación	2
4.4 Subproceso de Funciones Especiales.....	2
4.4.1 Factores de Expansión	2
4.4.2 Funciones de Agregación	2
4.4.3 Funciones de Tabulación.....	2
5. COMPONENTES DEL SISTEMA	2
5.1 Casos de Uso del Sistema Funcional de Procesamiento	2
5.2 Sistema Administrador de Bases de Datos	2
5.3 Sistema de Producción de Datos	2
6. METODOLOGÍA	2
PARA EL DISEÑO	2
PARA EL DISEÑO Y PRODUCCIÓN	2
PARA PRODUCCIÓN	2
7. DEFINICIÓN DE RIESGOS	2

	<p align="center">PROCESAMIENTO DE DATOS MICROESTABLECIMIENTOS - MICRO</p>	<p>CÓDIGO: SI-MICRO-MOT-02 VERSIÓN: 2 PAGINA: 3 FECHA: 30-04-09</p>
---	---	---

INTRODUCCIÓN

El presente documento tiene por objeto servir de guía, al staff informático, para diseñar e implementar el sistema de procesamiento de datos, en el cual se muestra la estructura y las características esenciales del sistema lo mismo que los lineamientos para las especificaciones de automatización de las diferentes operaciones estadísticas (OE). De igual manera el documento tiene como fin servir de instrumento de comunicación con los especialistas temáticos.

La consideración esencial para el diseño y desarrollo de un sistema de procesamiento moderno responde a las nuevas políticas del DANE respecto al mejoramiento en la productividad en el tratamiento informático de los datos de las diferentes operaciones estadísticas que de manera periódica y no periódica debe desarrollar la institución para responden a las crecientes necesidades de información básica.

El diseño y desarrollo de un nuevo sistema de procesamiento para el DANE responde a la creciente demanda de información por parte del estado colombiano y los diversos tipos de investigadores y usuarios tanto internos como externos para la planificación, toma de decisiones e investigación en los sectores: demográfico, económico, social, ambiental, etc. Estas amplias y diversas necesidades de información requieren de un sistema estadístico que responda de manera eficiente y con alta productividad.

En la actualidad la productividad en la institución no cumple con las necesidades de información básica que requiere el país; debido en parte a que existe una proliferación de formas y formatos que se tienen para su almacenamiento, análisis y administración de los datos; realizándose estos procesos de manera independiente, con distintos procedimientos y herramienta en las diferentes secciones encargadas de las operaciones estadísticas (OE). Conllevando a que la accesibilidad no sea eficiente por cuanto solamente los directamente involucrados con la OE saben donde se encuentran los datos, conllevando a que se dupliquen archivos y se vuelva más costosa e ineficiente su explotación.

Con el nuevo sistema se pretende unificar la forma de obtención, revisión, codificación, validación, procesamiento, consolidación y difusión de la información, atendiendo a la implantación de procesos y procedimientos automatizados eficaces a través del uso adecuado de la tecnología de la información TI.

1. SISTEMA DE PROCESAMIENTO DE DATOS

El sistema de procesamiento de datos del DANE se encuentra enmarcado dentro de un suprasistema que esta compuesto por un sistema de recolección, un sistema de procesamiento y un sistema de difusión de datos; como se observa en la figura No. 1.

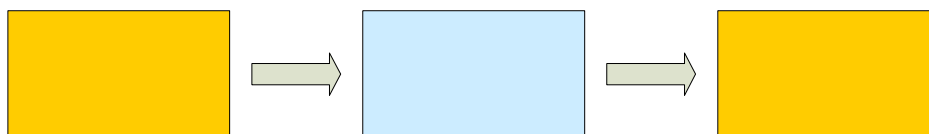


Figura No. 1. Subsistemas de un sistema estadístico típico.

Según la figura, el sistema de procesamiento lo podemos definir como un macroproceso el cual recibe insumos (datos) del sistema de recolección, los adecua y transforma convirtiéndolos en datos estadísticos para entregárselos al sistema de difusión, quien es el encargado de difundirlos entre los diferentes usuarios nacionales e internacionales.

Recolección de datos

Procesamiento de datos

2. OBJETIVOS

2.1 Objetivo General

El objetivo general frente al sistema de procesamiento es el de diseñar y poner en marcha un sistema automatizado que responda eficientemente al control de los datos recolectados de las diferentes Operaciones Estadísticas OE, su tratamiento y la disponibilidad de la información estadística con oportunidad, claridad, confiabilidad y accesibilidad a los diferentes usuarios internos.

2.2 Objetivos Específicos

- Desarrollar e implementar un sistema para controlar la información suministrada por el sistema de recolección.
- Desarrollar e implementar un sistema para el cargue de los datos, en estructuras unificadas.
- Desarrollar e implementar un sistema de depuración e imputación de datos para las diferentes OE que procesa la entidad.
- Desarrollar e implementar un sistema de administración de las bases de datos del sistema.
- Diseñar un ambiente para el análisis de datos.
- Diseñar un ambiente para la producción de los datos.

3. SERVICIOS DE INFORMACIÓN

El sistema a desarrollar deberá servir para de manera unificada y estandarizada transformar los datos obtenidos para las diferentes OE convirtiéndolos en información estadística de calidad, cuyas características frente al usuario interno están determinadas por su oportunidad, claridad, confiabilidad y accesibilidad.

Los datos e información que debe estar disponible en el sistema a los usuarios son: los **microdatos** no identificables libres de inconsistencias de las diferentes operaciones estadísticas que maneja el DANE; los **macrodatos** generados por operaciones realizadas sobre los microdatos, dichos macrodatos deberán estar disponibles en tablas de bases de datos o en forma de tabulados; y, los **metadatos** indispensables para facilitar la utilización de los microdatos y macrodatos.

La principal característica de los datos disponibles es que debe ser fácilmente accesible por los usuarios internos para ser tratada y analizada para la generación de nueva información y conocimiento.

4. MODELO FUNCIONAL DEL SISTEMA

A través del diagrama de flujo de datos DFD de la figura No.2 se muestra el modelo funcional general del sistema, con sus principales procesos, que a continuación se describen de manera general y se determinan algunas de sus restricciones. Es fundamental entender que algunos procesos se encuentran implícitos, por cuanto están dentro de la lógica de este tipo de aplicaciones informáticas, tal es el caso de los procesos de análisis que se realizan a través de todos los procesos.

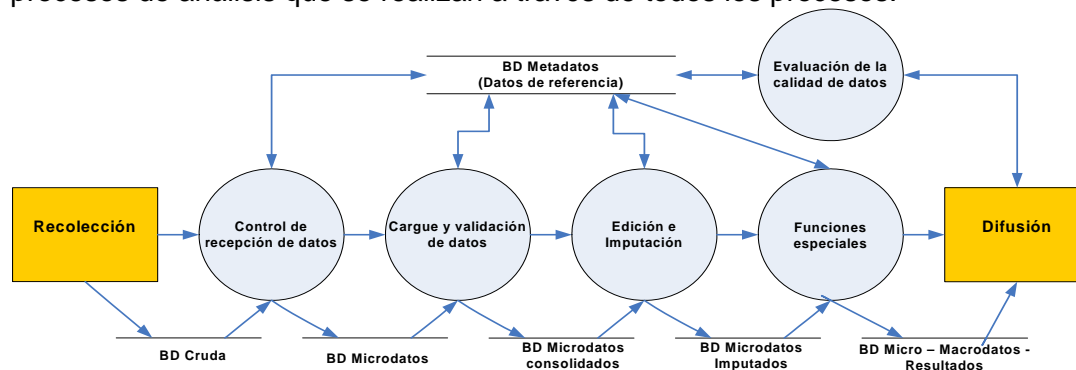


Figura No. 2. Diagrama de Flujo de Datos del Sistema de Procesamiento

4.1 Subproceso de Control de Recepción de Datos

Tan pronto llegan los datos crudos de recolección el proceso se encargará de verificar entre otros: que los datos estén completos de acuerdo con el diseño estadístico, v. g. que no hagan falta datos de las áreas geográficas provistas para recolección; que los archivos y registros vengan dentro de las estructuras de datos definidas; que los registros traigan los ítems de identificación necesarios y suficientes y que no vengan registros o archivos duplicados.

	<p align="center">PROCESAMIENTO DE DATOS MICROESTABLECIMIENTOS - MICRO</p>	<p>CÓDIGO: SI-MICRO-MOT-02 VERSIÓN: 2 PAGINA: 6 FECHA: 30-04-09</p>
---	---	---

Como salidas del proceso se suministrará al proceso de cargue una base de datos completa con los ítems de identificación completos y reenumerados si es necesario, complementando los ítems con base en los datos de referencia de la base de metadatos cuando sea indispensable (ver figura No. 2).

4.2 Subproceso de Cargue y Validación de Datos

Una condición fundamental para facilitar el procesamiento automatizado es la estructuración formal de los datos y la utilización de las herramientas de software apropiadas. Con estas condiciones se cargan los datos a la Base de Datos y se adecuan pasando de una estructura vertical a una estructura horizontal y controlando la integridad y duplicidad de los datos si es necesario.

Una de las políticas de la entidad determina que el administrador de bases de datos ORACLE debe ser el repositorio de microdatos de las operaciones estadísticas, por tal motivo se deben diseñar las estructuras para este manejador de tal manera que se facilite su depuración por los informáticos y el análisis por los especialistas temáticos.

Por otra parte, se conoce que existe software especializado en procesamiento de datos estadísticos como el CPro¹, pero su versión original trabaja con archivos planos, lo cual quiere decir que hay necesidad de generar otras estructuras de datos y/o interfases para la depuración de los datos. Adicionalmente, para la revisión de información se cuenta con el paquete estadístico SAS.

Para realizar el procesamiento de cada trimestre de la encuesta de Microestablecimientos, se deben realizar los siguientes pasos:

1. Se realiza la actualización de campos en la tabla de homologación de la encuesta para el nuevo periodo de la investigación:

En la tabla GEA_HMLGACION_PRGNTAS

Los campos con la nueva información son:

NRO_ENCSTA = numero de la encuesta para el trimestre.

VRSION_ENCSTA= numero de la versión de la encuesta para el trimestre.

¹ El CPro (**C**ensus and **S**urvey y **P**rocessing System) es un paquete de software para entrar, editar, tabular y diseminar datos de censos y encuestas. CPro combina las características del IMPS (**I**ntegrated **M**icrocomputer **P**rocessing System) y el ISSA (**I**ntegrated **S**ytem for **S**urvey **A**nalysis) y es desarrollado por la oficina de Censos de los Estados Unidos de América.

```

INSERT INTO DANE_MICRO.GEA_HMLGCION_PRGNTAS (
    CDGO_PRGNTA, CDGO_SBPRGNTA, BLOQUE,
    CDGO_CAMPO, ORDEN_CMPO, NOMBRE_TABLA,
    NRO_ENCSTA, VRSION_ENCSTA, ESTDO_CAMPO,
    ESTDO_RVSION, RGSTROS_NO_NULO, RGSTROS_RSPSTAS,
    RANGO_VLDCION, TIPO_VLDCION, ERROR_RANGO,
    TIPO_DATO, NMBRE_ESQMA, CDGO_ENCSTA,
    TXTO_PRGNTA)
select    CDGO_PRGNTA, CDGO_SBPRGNTA, BLOQUE,
          CDGO_CAMPO, ORDEN_CMPO, NOMBRE_TABLA,
          NRO_ENCSTA, VRSION_ENCSTA, ESTDO_CAMPO,
          ESTDO_RVSION, RGSTROS_NO_NULO, RGSTROS_RSPSTAS,
          RANGO_VLDCION, TIPO_VLDCION, ERROR_RANGO,
          TIPO_DATO, NMBRE_ESQMA, CDGO_ENCSTA,
          TXTO_PRGNTA
from      DANE_MICRO.GEA_HMLGCION_PRGNTAS
where     nro_encsta = 2402 -- numero de la encuesta anterior
and      vrsion_encsta = 3 -- numero de la versión anterior de
la encuesta

```

2. Para identificar las nuevas preguntas se realiza el siguiente sql

```

insert into GEA_HMLGCION_PRGNTAS
with preguntas as
(select    d.NRO_PREGUNTA, d.DESCRIPCION, d.LISTA_VALOR, d.TIPO_VALOR,
          h.ORDEN, h.BLOQUE, h.VERSION, h.NRO_ENCUESTA
from      TGEA_preguntas d join TGEA_PREGUNTAS_ENC h
on        (d.nro_pregunta = h.NRO_pregunta)
where     h.NRO_ENCUESTA = 2402
AND      h.VERSION = 3),
datos1
as
(select    v.* , m.NRO_SUBPREGUNTA , m.descripcionS
from      preguntas v left join tgea_subpreguntas m
on        (v.NRO_PREGUNTA = m.NRO_PREGUNTA))
select f.NRO_PREGUNTA, f.nro_subpregunta, f.BLOQUE, 'nombre
columna' || f.NRO_PREGUNTA || f.nro_subpregunta , f.ORDEN,
'MIC_FORMULARIO_INICIAL' ,
f.NRO_ENCUESTA, f.VERSION, 'AC' , '2', null, null,
null,null,null, decode(f.tipo_valor, 'N' , 'NUMBER' , 'VARCHAR2') ,
'DANE_MICRO', 'MEST',
f.NRO_PREGUNTA || ' --' || f.descripcion || ' ' || f.descripcionS
from      datos1 f left join GEA_HMLGCION_PRGNTAS g
on        (f.nro_pregunta = g.cdgo_prgnta and f.bloque = g.bloque
and nro_encsta = 2402 and vrsion_encsta = 4 )
where     g.CDGO_CAMPO is null

```

	PROCESAMIENTO DE DATOS MICROESTABLECIMIENTOS - MICRO	CÓDIGO: SI-MICRO-MOT-02 VERSIÓN: 2 PAGINA: 8 FECHA: 30-04-09
---	---	---

el campo '**nombre columna**' || f.NRO_PREGUNTA || f.nro_subpregunta de debe
reemplazar por el nombre de la columna nueva para la tabla
GEA_HMLGCIION_PRGNTAS

3. Con la siguiente consulta se realiza la actualización de las tablas

mic_formulario_inicial
mic_formulario_inicial_inc
mic_formulario_inicial_final

```
select rownum, 'alter table ' || nombre_tabla || ' add (' || cdgo_campo || '
' || tipo_dato || '(20));'
from GEA_HMLGCIION_PRGNTAS
where nro_encsta = 2402 --numero de la encuesta
and vrsion_encsta = 3 --versión de la encuesta
and cdgo_prgnta = 1634 -- código de la pregunta
```

para poder generar los job en la base de datos se debe actualizar la tabla
GEA_LANZA_PROCESO con la información de la encuesta a trabajar según ejemplo

```
INSERT INTO DANE_MICRO.GEA_LANZA_PROCESO (
    NMRO_ENCSTA, VRSION_ENCSTA, ANO_ENCSTA,
    PRDO_ENCSTA, SIGLA_PROCESO, IND_PRCISO,
    DSCRPCION_PRCISO)
VALUES (2402, 3 , 2009 , 1 , 'ENC' , 'S' , 'Llena encuesta Micro' )
```

4. Con las tablas actualizadas se generan los job en la base de datos Oracle para
realizar esta tarea de transposición de datos de la tabla tgea_respuestas a
mic_enlistamineto_inicila y mic_formulario_inicial según correspondan en el
enlistamiento o la encuesta.

```
DANE_MICRO.PR_LANZAR_LLENA_ENCUESTA
('MEST_II' /* VARCHAR2 */ nombre de la encuesta,
2401 /* BINARY_INTEGER */ numero encuesta,
2 /* BINARY_INTEGER */ versión,
2009 /* BINARY_INTEGER */ año de referencia,
1 /* BINARY_INTEGER */ periodo de la encuesta,
3 /* BINARY_INTEGER */ estado ,
1500 /* BINARY_INTEGER */ ); cantidad de registro a procesar
```

```
DANE_MICRO.PR_LANZAR_LLENA_ENCUESTA
('MEST_II' /* VARCHAR2 */ , nombre de la encuesta
2402 /* BINARY_INTEGER */ , numero encuesta,
3 /* BINARY_INTEGER */ , versión
2009 /* BINARY_INTEGER */ , año de referencia
```


	PROCESAMIENTO DE DATOS MICROESTABLECIMIENTOS - MICRO	CÓDIGO: SI-MICRO-MOT-02 VERSIÓN: 2 PAGINA: 9 FECHA: 30-04-09
---	---	---

1 /* BINARY_INTEGER */ , periodo de la encuesta
3 /* BINARY_INTEGER */ , estado
1500 /* BINARY_INTEGER */); cantidad de registro a procesar

DANE_MICRO.PR_LANZAR_LLENA_ENCUESTA_INC
('MEST_II' /* VARCHAR2 */ , nombre de la encuesta
2402 /* BINARY_INTEGER */ , numero encuesta
3 /* BINARY_INTEGER */ , versión
2009 /* BINARY_INTEGER */ , año de referencia
1 /* BINARY_INTEGER */ , periodo de la encuesta
2 /* BINARY_INTEGER */ , estado
500 /* BINARY_INTEGER */); cantidad de registro a procesar

Los job se ejecutan con intervalos de tiempo solicitados por los usuarios para cada investigación e invocan al siguiente procedimiento:

```
PR_LANZAR_LLENA_ENCUESTA(
    minv_enc varchar2,      -- codigo de la encuesta
    mnum_enc pls_integer,   -- numero de la encuesta
    mver_enc pls_integer,   -- version de la encuesta
    mano_ref pls_integer,   -- año de la encuesta
    mper_ref pls_integer,   -- periodo de la encuesta
    mest_enc pls_integer,   -- estado del registro de la encuesta
    mreg_pro pls_integer    -- numero de registros procesados
)
```

5. El procedimiento anterior realiza el control para la transposición de los datos y el llamado al paquete PKG_GEA_PROCESO el cual realiza las tareas propias de transposición de datos a las tablas mic_enlistamiento_inicial y mic_formulario_inicial y controlando la integridad y duplicidad de los datos si es necesario.

Con el siguiente job se realiza los cálculos de variables

DANE_MICRO.PR_ACT_RESULTADO_ENC2008
(1 /* BINARY_INTEGER */ ,
2009 /* BINARY_INTEGER */ ,
1 /* BINARY_INTEGER */ ,
1 /* BINARY_INTEGER */);

El job se ejecuta con intervalos de tiempo solicitados por los usuarios y se invoca el procedimiento:

`pr_act_resultado_enc2008`

```

(mtrim      pls_integer,  --- trimestre a procesar
mano        pls_integer,  --- año del procesamiento
mestruct   pls_integer,  --- estructura 1
mtics       pls_integer)  --- proceso tics 1

```

6. Después de realizar la recolección del enlistamiento se procede a generar la muestra de la encuesta a realizar por departamento y municipio. Para realizar dicha muestra se debe correr el siguiente procedimiento

```

pr_act_superv_usm
(mnum_enc pls_integer,  -- número de la encuesta
mver_enc pls_integer,  -- versión de la encuesta
mano_ref pls_integer,  -- año de la encuesta
mper_ref pls_integer  -- periodo de la encuesta
)

```

7. Para realizar la validación y consistencia de la información de las encuestas recolectadas se actualiza la tabla GEA_VLDCION_PRGNTA con la información Excel entregada por los encargados de la encuesta. Después de realizar la actualización de la tabla y de verificar que las sentencias sql estén bien escritas, se realiza el siguiente sql para el llamado a los procedimientos.

```

-- COPIAR MIC_FORMULARIO_INICIAL EN MIC_FORMULARIO_FINAL
select count(1) from MIC_FORMULARIO_inicial a where A.PRDO_RFRNCIA = 4
AND A.ANO_RFRNCIA=2008;
select count(1) from MIC_FORMULARIO_inicial_INC a where A.PRDO_RFRNCIA
= 4 AND A.ANO_RFRNCIA=2008;
select count(1) from MIC_FORMULARIO_FINAL a where A.PRDO_RFRNCIA = 4
AND A.ANO_RFRNCIA=2008;

--INSERT INTO MIC_FORMULARIO_FINAL SELECT * FROM
MIC_FORMULARIO_final_evo A WHERE A.PRDO_RFRNCIA = 1 AND
A.ANO_RFRNCIA=2008;
INSERT INTO MIC_FORMULARIO_FINAL
SELECT * FROM MIC_FORMULARIO_INICIAL A
WHERE A.PRDO_RFRNCIA = 4 AND A.ANO_RFRNCIA=2008 and
a.NRO_ENCUESTA=2402;
union
INSERT INTO MIC_FORMULARIO_FINAL
SELECT * FROM MIC_FORMULARIO_INICIAL_INC A
WHERE A.PRDO_RFRNCIA = 2 AND A.ANO_RFRNCIA=2008 and
a.NRO_ENCUESTA=2402;
COMMIT;

select a.RSLTDO_ENTRVSTA_TMTCO_EV
from mic_formulario_final a

```

	PROCESAMIENTO DE DATOS MICROESTABLECIMIENTOS - MICRO	CÓDIGO: SI-MICRO-MOT-02 VERSIÓN: 2 PAGINA: 11 FECHA: 30-04-09
---	---	--

```
where A.PRDO_RFRNCIA = 4 AND A.ANO_RFRNCIA=2008 and
a.NRO_ENCUESTA=2402
```

```
update mic_formulario_FINAL a
set a.RSLTDO_ENTRVSTA_TMTCO_EV=a.RSLTDO_ENTRVSTA_EVLCION --LA VARIABLE
ES NULA
where A.PRDO_RFRNCIA = 4 AND A.ANO_RFRNCIA=2008 and
a.NRO_ENCUESTA=2402;
COMMIT;
```

```
-- PROCESO DE CONTEO DE INCONSISTENCIAS TRIMESTRAL inicial
```

```
-- Num_encuesta, Per_encuesta, Año encuesta, frecuencia
exec pr_lanza_valid4( 2102, 1, 2008, 3);
```

```
-- PROCESO DE CONTEO DE INCONSISTENCIAS ANUAL INICIAL
-- Num_encuesta, Per_encuesta, Año encuesta, frecuencia
exec pr_lanza_valid4 ( 2402, 4, 2008, 12);
```

```
-- actualizacion de errores inicial
update gea_vldcion_prnta a set a.NMRO_ERRRES_INI = a.NMRO_ERRRES_FIN
where a.NMRO_ENCSTA = 2402
and a.ANO_RFRNCIA=2008
and a.PRDO_RFRNCIA=4;
commit;
```

```
-- PROCESO DE CORRECCION DE INCONSISTENCIAS
--- Parametro frecuencia de la inconsistencia
---- freq (3:evolucion 12:estructura), trm, año
exec pr_lanza_correccion ( 3, 1, 2008 );
```

```
exec pr_lanza_valid4( 2102, 1, 2008, 3);
exec pr_correccion ( 12 );
```

```
--- Parametro frecuencia de la inconsistencia Estructura
---- freq (3:evolucion 12:estructura), trm, año,nro encuesta
exec pr_lanza_correccion ( 3, 4, 2008, 2402 );
```

```
-- Num_encuesta, Per_encuesta, Año encuesta, frecuencia
exec pr_lanza_valid4(2402, 4, 2008, 3);
exec pr_correccion ( 12 );
```

	<p align="center">PROCESAMIENTO DE DATOS MICROESTABLECIMIENTOS - MICRO</p>	<p>CÓDIGO: SI-MICRO-MOT-02 VERSIÓN: 2 PAGINA: 12 FECHA: 30-04-09</p>
---	---	--

4.3 Subproceso de Edición e Imputación

Este proceso es uno de los más relevantes en lo que concierne a la calidad de los datos estadísticos (no se aplica a todas las investigaciones) y sus funciones principales son: verificar y controlar la integridad de las secciones de los datos (registros) que integran una unidad de observación estadística; y, corregir con base en metodologías y normas de edición e imputación, definidas por los especialistas temáticos, los valores faltantes, inválidos o inconsistentes.

Un sistema de edición e imputación puede contener más de un programa de computador. Es usual al menos tener dos programas para el sistema de edición e imputación: el primero es usado para encontrar y reportar los errores de estructura los cuales deben ser resueltos antes de tratar los datos en la próxima etapa. Estos incluyen registros missing, registros duplicados u otras incompatibilidades con los requerimientos los cuales son considerados esenciales para el procesamiento. El segundo programa es el de edición tradicional el cual valida y hace consistencia de los valores de los datos.

La interacción de los especialistas temáticos con los informáticos en este proceso es significativamente alta por cuanto las combinaciones de las interrelaciones entre los valores de las diferentes variables crece geométricamente y no se pueden prever con anticipación todas las posibles inconsistencias que se pueden presentar.

La utilización de software especializado en procesamiento de datos estadísticos facilita el control y seguimiento de las especificaciones implementadas para la depuración de los datos porque es más fácil de programar y permite realizar rastreos de lo realizado e ir obteniendo estadísticas de imputación de manera automática. En la actualidad para el procesamiento estadístico contamos con el paquete CPro, software que reemplazó al IMPS.

4.4 Subproceso de Funciones Especiales

Aunque lo fundamental en el procesamiento de datos es la obtención de datos limpios y consistentes, existen un conjunto de funciones posteriores y que complementan la adecuación y preparación de los datos para ser usados tanto por especialistas como por usuarios corrientes, estos subprocesos se describen a continuación y su precedencia se pueden observar en la figura No. 3.

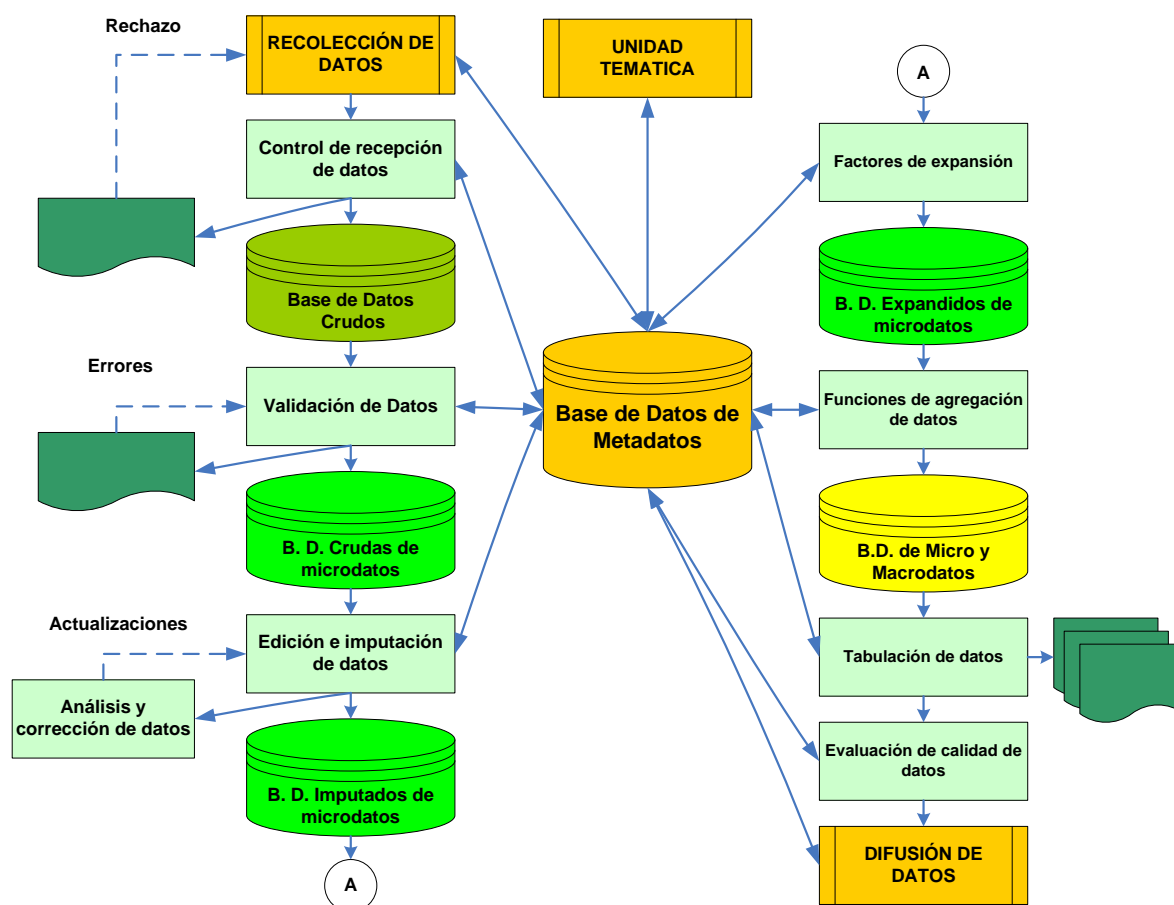


Figura No. 3. Flujo del Procesamiento de Datos

4.4.1 Factores de Expansión

Como por lo general se trabaja con muestras, obliga a calcular previamente, unos ciertos parámetros, factores de elevación necesarios para obtener las estimaciones. Este subproceso se realiza después de obtener los datos de las OE totalmente depurados como se puede observar en la figura No. 3. Este subproceso en algunos casos puede estar incluido en el subproceso de tabulación.

4.4.2 Funciones de Agregación

Las funciones de agregación también deben realizarse con datos depurados o limpios, conformado de esta manera los denominados macrodatos, como se puede observa en la figura No.3. En OE continuas como son la mayoría de las que trata el DANE se van

	<p align="center">PROCESAMIENTO DE DATOS MICROESTABLECIMIENTOS - MICRO</p>	<p>CÓDIGO: SI-MICRO-MOT-02 VERSIÓN: 2 PAGINA: 14 FECHA: 30-04-09</p>
---	---	--

consolidando las series de tiempo, que deberán ser empalmadas por ejemplo cuando se cambia de base en la operación Índice de Precios al Consumidor (IPC).

4.4.3 Funciones de Tabulación

Al proceso de tabulación suelen entrar microdatos obteniéndose como salida macrodatos, bien sean tablas, parámetros estadísticos como medidas de posición, dispersión, correlación etc. que permiten deducir el comportamiento de ciertos fenómenos o hacer predicciones etc. Pero no solamente al proceso de tabulación entran microdatos también pueden entrar macrodatos y si provienen de estadísticas continuas generar predicciones o pronósticos de algún fenómeno a través del análisis de series de tiempo.

5. COMPONENTES DEL SISTEMA

En esta parte del documento se encuentra de una forma sistemática y complementada la descripción de la funcionalidad del sistema a través de “casos de uso”²; y para completar los componentes del sistema se describen los componentes principales de apoyo, que hacen parte de la *arquitectura del sistema* (ver figura No. 4).

El componente fundamental que hace parte de la infraestructura del sistema de procesamiento es el Sistema de Base de Datos, el cual es el encargado de a) administrar de los datos (entendiéndose como los microdatos, macrodatos y metadatos), b) administrar las aplicaciones de tratamiento estadístico y el código reutilizable, y c) servir de soporte para la seguridad, integridad y confidencialidad de los datos.

² Un caso de uso es una descripción de un conjunto de secuencia de acciones, incluyendo variantes, que ejecuta un sistema para producir un resultado observable de valor para una actor. (Booch Grady y otros. El Lenguaje Unificado de Modelado. Addison Wesley. España. 1999.

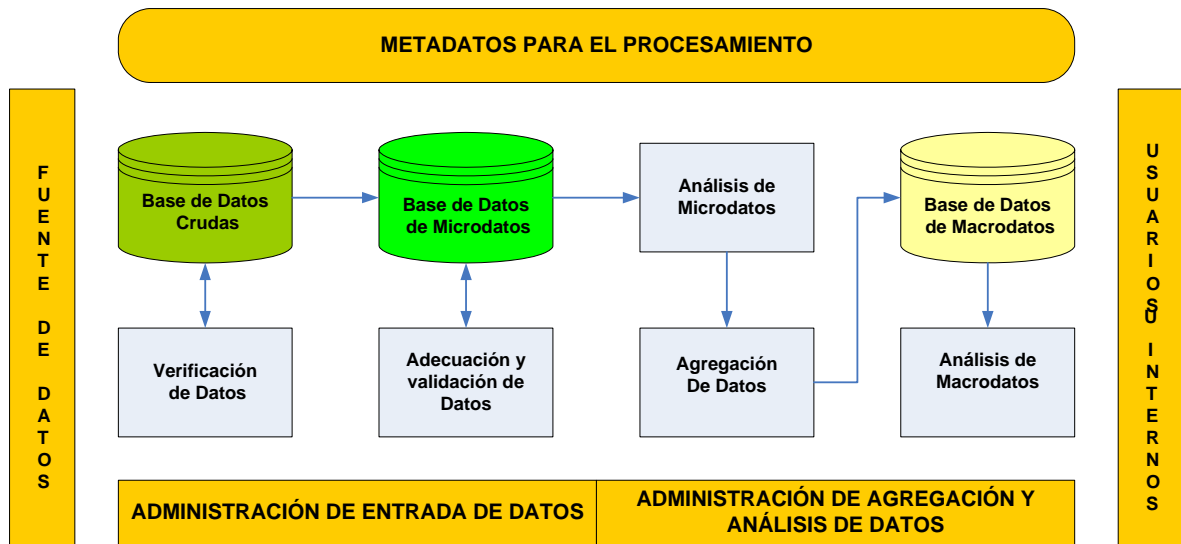


Figura No. 4. Arquitectura del Sistema de Procesamiento

	PROCESAMIENTO DE DATOS MICROESTABLECIMIENTOS - MICRO	CÓDIGO: SI-MICRO-MOT-02 VERSIÓN: 2 PAGINA: 16 FECHA: 30-04-09
---	---	--

5.1 Casos de Uso del Sistema Funcional de Procesamiento

Caso de Uso	Control de Recepción de Datos de las OE
Descripción	Recepción de los datos de cada OE del DANE, que se han especificado para ser recogidos, organizados y entregados por recolección.
Actores	Recolección de datos Técnico en sistema
Pre-condiciones	Datos organizados por lotes bien delimitados para facilitar el control, por ejemplo por áreas geográficas o por municipio.
Post-condiciones	Datos completos de las OE con sus adecuados reportes para construir la metadata del procesamiento.
Acciones	Repetir control hasta cuando los datos por OE estén completos: <ul style="list-style-type: none"> • Verificar que los datos de recolección vayan organizados dentro de las estructuras definidas, completos, consistentes y de acuerdo al reporte de soporte. • Verificar y controlar que los ítems de identificación vayan completos. • Verificar y controlar la duplicidad de archivos y registros de datos de las diferentes OE. Fin de repetición. Identificar que hizo falta y generar reporte de soporte.
Variaciones	
Requerimientos Especiales	
Cuestiones	Cuando los datos no son capturados por DMC y se estén

	PROCESAMIENTO DE DATOS MICROESTABLECIMIENTOS - MICRO	CÓDIGO: SI-MICRO-MOT-02 VERSIÓN: 2 PAGINA: 17 FECHA: 30-04-09
---	---	--

	asignando códigos directamente por la máquina, se deben generar programas para codificación o realizar codificación manual.
--	---

Caso de Uso	Cargue de Datos
Descripción	Se pretende cargar los datos de las OE en estructuras formales para ser tratados, depurados y utilizados por los distintos tipos de usuarios internos.
Actores	Técnico en sistemas Especialistas temáticos
Pre-condiciones	Los datos de una OE deberán estar completos
Post-condiciones	Cuando se realiza edición e imputación de manera automática con software que requiera archivos planos estos deberán suministrarse en estos formatos.
Acciones	<ul style="list-style-type: none"> • Renumerar los ítems de identificación, cuando sea requerido. • Incluir ítems de identificación que se encuentran en archivos de referencia. • Adecuar los datos originales (crudos) en estructuras de microdatos en el manejador ORACLE. • Generación de archivos planos en estructura de datos adecuadas para la edición e imputación (la imputación si se requiere dependiendo de la investigación). • Verificar y corregir integridad de registros de las unidades de observación. • Generación de archivos para codificación asistida por computador o manual. • Generación de archivos para cálculo de factores de expansión. • Generación de archivos para arreglo de variables que integran las unidades de observación. • Consolidar los archivos que se han seleccionado de las

	PROCESAMIENTO DE DATOS MICROESTABLECIMIENTOS - MICRO	CÓDIGO: SI-MICRO-MOT-02 VERSIÓN: 2 PAGINA: 18 FECHA: 30-04-09
---	---	--

	bases de datos originales para arreglos de unidades de información, imputación y realización de otras funciones de tratamiento estadístico.
Variaciones	Cuando no se generan nuevos archivos para el tratamiento estadístico, es necesario generar interfaces que permitan hacer transparente el procesamiento estadístico.
Requerimientos Especiales	
Cuestiones	En caso de generar interfaces hay que considerar el desarrollo o adecuación de estas.

Caso de Uso	Edición e Imputación (E&I)
Descripción	La edición e imputación deberá dejar los microdatos con integridad entre las unidades de observación y los valores omitidos, inválidos o inconsistentes corregidos.
Actores	Especialistas en informática Especialistas temáticos
Pre-condiciones	Las bases de datos o archivos generados deberán poseer todos los ítems de identificación correctos.
Post-condiciones	Los datos deberán quedar consistentes y validados por los especialistas temáticos. Todas las normas implementadas deberán quedar probadas y documentadas.

	PROCESAMIENTO DE DATOS MICROESTABLECIMIENTOS - MICRO	CÓDIGO: SI-MICRO-MOT-02 VERSIÓN: 2 PAGINA: 19 FECHA: 30-04-09
---	---	--

Acciones	<p>Para cada OE, se debe controlar, verificar y corregir la integridad de los datos por unidad de observación³. Explorar los datos por unidades de observación y por variables conjuntamente con sus interrelaciones, antes de imputar para observar el estado inicial de la calidad de los datos. Hacer hasta que los datos queden sin errores:</p> <ul style="list-style-type: none"> • Verificar de manera automática la consistencia de los datos de acuerdo con las especificaciones, haciendo visibles los errores. • Corregir inconsistencias en forma asistida por computador. • Corregir las inconsistencias de manera automática, tales como valores faltantes, inválidos e inconsistentes. • Analizar los datos a través de procedimientos estadísticos para comprobar que las distribuciones de los datos no han cambiado. • Generar nuevas normas, de ser necesarias, corregir los programas y actualizar documentación. <p>Fin de Hacer. Generar bitácora de errores, estadísticas de imputación (si se requiere) y actualización de metadatos para usuario final.</p>
Variaciones	<p>Cuando existen algunos errores que se salen de lo común, estos se corrigen con programas especiales de computador.</p>
Requerimientos Especiales	<p>Entre los requerimientos especiales esta la medición de la calidad y el soporte para los procesos de validación.</p>
Cuestiones	

Caso de Uso	Factores de Expansión
Descripción	Obtener los factores de expansión para las variables de las OE que

³ Una unidad de observación es el equivalente a un cuestionario (Questionnaire en los paquetes de procesamiento estadístico) y puede estar integrada por uno o más tipos de registros interrelacionados que conforman una unidad.

	PROCESAMIENTO DE DATOS MICROESTABLECIMIENTOS - MICRO	CÓDIGO: SI-MICRO-MOT-02 VERSIÓN: 2 PAGINA: 20 FECHA: 30-04-09
---	---	--

	se realizan por muestra.
Actores	Técnico sistemas Estadístico Especialista temático
Pre-condiciones	Esta actividad se lleva a cabo con archivos depurados e información completa de las OE
Post-condiciones	Archivos de datos con sus factores de expansión incluidos.
Acciones	<ul style="list-style-type: none"> • Análisis de los datos depurados de las OE • Cálculo de factores de expansión • Realización de pruebas con datos expandidos.
Variaciones	No se calculan factores de expansión cuando las OE son de tipo censo.
Requerimientos Especiales	
Cuestiones	

Caso de Uso	Funciones de Agregación
Descripción	Realizar funciones de agregación sobre los microdatos generando macrodatos y funciones de agregación sobre los mismos macrodatos para y mantenerlos en tablas de la base de datos.
Actores	Técnico en sistemas Especialista temático

	PROCESAMIENTO DE DATOS MICROESTABLECIMIENTOS - MICRO	CÓDIGO: SI-MICRO-MOT-02 VERSIÓN: 2 PAGINA: 21 FECHA: 30-04-09
---	---	--

Pre-condiciones	Se debe partir de microdatos totalmente depurados
Post-condiciones	Los macrodatos deben quedar almacenados en la base de datos.
Acciones	<ul style="list-style-type: none"> • Generar macrodatos de microdatos o macrodatos poblando las tablas correspondientes • Realizar análisis a través de series de tiempo.
Variaciones	Cuando las OE no son continuas por lo general no se guardan los macrodatos en tablas de bases de datos, se guardan en forma de tabulados.
Requerimientos Especiales	De acuerdo a cada OE.
Cuestiones	

Caso de Uso	Funciones de tabulación
Descripción	Generar tabulados para análisis con base en microdatos y macrodatos (estructurados en bases de datos), con o sin factores de expansión.
Actores	Técnico en sistemas Especialista temático
Pre-condiciones	Poseer archivos en estructuras de datos adecuadas.
Post-	Tabulados validados por los especialistas temáticos con

condiciones	expansiones o sin ellas, según el caso.
Acciones	<ul style="list-style-type: none"> • Adecuar programas de tabulados según solicitud. • Generar tabulados con microdatos. • Generar tabulados con macrodatos (estructurados) • Analizar resultados de tabulados.
Variaciones	Los tabulados con datos sin depurar se utilizan para verificar las inconsistencias.
Requerimientos Especiales	
Cuestiones	

5.2 Sistema Administrador de Bases de Datos

Este sistema es el encargado de generar un ambiente adecuado para la administración de la información estadística por los “responsables de las operaciones estadísticas” o especialistas temáticos, y por otra parte, generar un ambiente efectivo para administrar y salvaguardar los datos de todas las operaciones estadísticas por parte de un administrador de base de datos⁴. Entre sus principales funciones que soporta el sistema de base de datos del DANE están las siguientes:

- Servir de herramienta de soporte para la administración y control de las bases de datos de microdatos, macrodatos y metadatos.
- Ser respaldo en la administración y control de las aplicaciones de tratamiento y análisis estadístico y utilización de código reutilizable.
- Facilitar la explotación y análisis de los datos a través de interfases de usuarios.
- Apoyar tecnológicamente la preservación de la seguridad, integridad, y confidencialidad de los datos.

5.3 Sistema de Producción de Datos

⁴ El administrador de la base de datos es la persona o grupo de personas encargadas de gestionar y controlar las actividades que tienen que ver con los datos de la entidad y con la base de datos respectivamente. Entre sus responsabilidades frente a los datos esta la de velar por la recuperabilidad, integridad, seguridad, disponibilidad, desempeño y desarrollo y el soporte a pruebas.

El sistema de producción de datos del Sistema de Procesamiento está supeditado a la cantidad de OE que maneja el DANE, su periodicidad y los subprocesos necesarios para cada operación estadística.

Paralelamente a la producción del procesamiento de datos, se debe realizar funciones de administración: de datos, aplicaciones y el código de reutilizable. De igual manera se debe ir actualizando la documentación.

6. METODOLOGÍA

El desarrollo del sistema de procesamiento esta enmarcado dentro del enfoque avanzado que se muestra en la figura No. 5. Contrario al enfoque tradicional, aplicado especialmente para sistemas simples donde se pueden desarrollar las fases una a continuación de la otra y los procesos en cascada, se diseña⁵ y preparan completamente las herramientas de software en la primera fase y luego se entra a la fase de producción. Si se aplica el enfoque tradicional para desarrollar sistemas complejos el desarrollo se presenta como se observa en la figura No. 5, presentándose retrasos en la programación y por consiguiente aumentándose los costos.

La condición necesaria para implantar el enfoque avanzado es el trabajo delimitado pero integrado entre especialistas temáticos, estadísticos y especialistas informáticos.

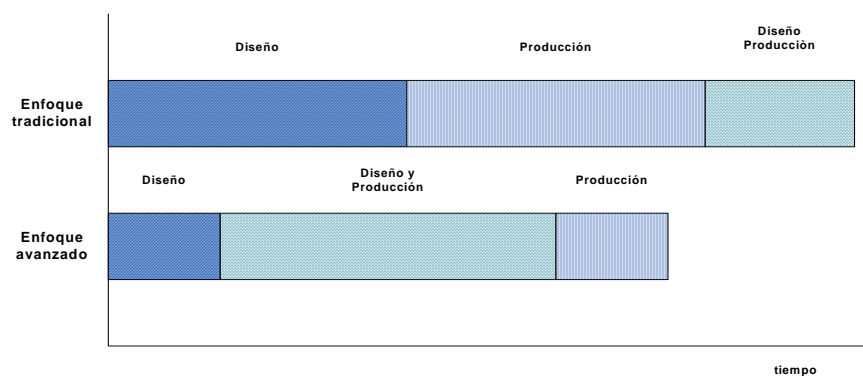


Figura No. 5. Enfoque para desarrollo de Sistemas Estadísticos (De Miguel Adoración)

En este orden de ideas, tomando como base el enfoque avanzado, con sus tres fases, la metodología la siguiente:

PARA EL DISEÑO

⁵ El diseño en los sistemas estadísticos está compuesto por la planeación, diseño y desarrollo de todos los procedimientos y procesos automatizados o no.

	<p align="center">PROCESAMIENTO DE DATOS MICROESTABLECIMIENTOS - MICRO</p>	<p>CÓDIGO: SI-MICRO-MOT-02 VERSIÓN: 2 PAGINA: 24 FECHA: 30-04-09</p>
---	---	--

Para empezar la planeación de la Edición e Imputación se debe tener el diseño del cuestionario finalizado y la descripción de los registros/archivos estar disponibles para la operación estadística que se toma como prototipo.

- Diseñar las estructuras de datos para el procesamiento y administración de los datos.
- Desarrollar la guía metodológica de Edición e Imputación para cada una de las operaciones estadísticas.
- Definir la arquitectura del sistema de procesamiento.
- Desarrollar las especificaciones detalladas para computador.
- Desarrollar los programas de computador.

PARA EL DISEÑO Y PRODUCCIÓN

En esta fase se debe evaluar y analizar de manera conjunta entre especialistas temáticos, estadísticos y especialistas informáticos que tan completas, pertinentes y claras han sido las especificaciones para desarrollar los procesos automatizados.

- Pruebas unitarias e integrales de los procesos automatizados implementados.
- Análisis de datos por especialistas temáticos y estadísticos.
- Actualizaciones y adecuaciones de programas.
- Evaluar la calidad de los datos.
- Adecuar documentación.

PARA PRODUCCIÓN

En esta fase debe estar todos los procedimientos y procesos automatizados o no totalmente probados para que los técnicos en sistemas ejecuten esta tarea y los especialistas temáticos realicen las validaciones.

7. DEFINICIÓN DE RIESGOS

Los riesgos aquí presentados para el sistema de procesamiento de datos son los riesgos de software identificados después de diseñar la arquitectura del sistema, los riesgos más específicos se definirán cuando se realice el diseño detallado de cada una de las operaciones estadísticas. De la misma manera algunos riesgos generales no solamente son válidos para este sistema son validos para los sistemas de recolección y difusión de los datos.

La clasificación de potenciales riesgos del sistema están enmarcados dentro de las fases de desarrollo de software:

Necesidades y Requerimientos	<ul style="list-style-type: none"> • Los requerimientos no son completos, correctos o/y claros. • Resultan nuevos requerimientos hasta cuando los productos (software) ya están desarrollados.
Planificación y especificaciones	<ul style="list-style-type: none"> • Se determina una planificación muy optimista e lugar de realista. • La planificación no es completa por no integrar todos los componentes del sistema. • La arquitectura del sistema no es completa. • Las especificaciones son insuficientes y no incluyen las tareas necesarias.
Contratación	<ul style="list-style-type: none"> • La contratación tarda más de lo esperado. • No se suministra al personal contratado la instrucción necesaria.
Diseño y desarrollo	<ul style="list-style-type: none"> • El diseño o desarrollo es difícil de llevar a cabo. • Los procesos que determinan el sistema no se encuentran bien definidos. • El diseño queda demasiado simple que no cubre cuestiones principales (especificaciones detalladas) con lo cual hay que diseñar en la marcha con los traumatismos que esto conlleva. • Las interfaces para los usuarios no quedan bien definidas o su desarrollo es inadecuado. • El personal de desarrollo necesita entrenamiento y adaptación de las nuevas herramientas. • El no trabajo integrado desfavorece la productividad en el diseño y desarrollo. • Las actividades de control de calidad son recortadas haciendo que se tenga que repetir el trabajo. • La falta de entusiasmo en la gestión de riesgos impide que se detecten los riesgos más importantes
Integración del sistema	<ul style="list-style-type: none"> • No se determina el ambiente de pruebas del software. • Los ciclos de análisis de datos son muy largos atrasando el desarrollo del sistema.

	<p>PROCESAMIENTO DE DATOS MICROESTABLECIMIENTOS - MICRO</p>	<p>CÓDIGO: SI-MICRO-MOT-02 VERSIÓN: 2 PAGINA: 26 FECHA: 30-04-09</p>
---	--	--