Aspectos metodológicos de la consolidación y el procesamiento de los formularios F1 y F2

#

[Presentación](#h.jyscj3u31xnd)

[Esquema general del proceso](#h.w1cw2ql5esb9)

[Transmisión](#h.rwmjw6y3yptl)

[Revisión](#h.kdfmyunfng4f)

[Consolidación](#h.hkrz9kdgcin7)

[Procesamiento](#h.rfwilw10r6u2)

[Preparación de los datos](#h.anjfb5vzebgh)

[Análisis exploratorio de los datos](#h.uumi04qd2fcd)

[Producción de tablas y agregados estadísticos](#h.nk9fmwk2e0ef)

#

# Presentación

Este documento tiene como finalidad describir los procesos que se realizan para la consolidación y procesamiento de la información recolectada sobre la producción y demanda de información en la Alcadía de Cali usando los formularios F1 y F2, que fueron diligenciados durante el operativo de campo realizado en el marco de la formulación del PET-Cali.

En la siguientes secciones se describe el esquema general del proceso transmisión, consolidación y procesamiento de los formularios diligenciados. Posteriormente se desarrolla cada uno de estos procesos con más detalle.

# Esquema general del proceso



# Transmisión

El diligenciamiento de los formularios se hace usando una versión en Excel denominada F1e y F2e de los formularios F1 y F2 de la metodología de formulación PET del DANE. Terminada cada jornada los archivos diligenciados son subidos al servicio de almacenamiento de archivos online, agrupándolos en carpetas, una por cada encuestador, siguiendo el esquema para nombrar los formularios descrito en la *Guía para el diligenciamiento de los formularios[[1]](#footnote-0)*.

# Revisión

El proceso de revisión empieza con la organización de los formularios diligenciados en carpetas por dependencias, y en cada dependencias por tipo de formulario (F1 para la oferta y F2 para la demanda), esto con el propósito de verificar la cobertura de la encuesta en las dependencias y hacer una estimación del volumen de producción y de demandas de la entidad. Posteriormente se revisa con el equipo de encuestadores, el supervisor y el coordinador del operativo de campo la completitud de cada encuesta, si efectivamente el proceso estadístico descrito es una operación estadística (OE) y si es posible fusionar procesos de producción descritos en varios formularios como una sola OE. De este proceso sale el inventario de los formularios que serán cargados a la base de datos.

La siguiente revisión consiste en verificar el correcto flujo de los formularios y la revisión de la sintaxis usada en el diligenciamiento sea la correcta, e.g. el formato de fechas, números y uso de caracteres especiales para listar, definición de variables. Una vez revisado un formulario se asigna un número consecutivo de cargue. Durante el diligenciamiento es posible que los campos dispuestos para listar las variables y los indicadores o agregados producidos sea insuficiente. Para ello se usa otro formulario con el mismo nombre de base y se añade un número consecutivo para su identificación. Para los formularios principales se numera usando un entero y para los formularios derivados se asigna la parte entera del formulario base y se añade una parte decimal que identifica al formulario derivado ( e.g. formulario base = 10, primer formulario derivado=10.1).

# Consolidación

Para el proceso de consolidación se diseñó una aplicación web (desarrollada en PHP y ) que tiene una base de datos relacional en MySQL que mapea la estructura de los formularios. La aplicación asiste en el proceso de cargar los archivos en Excel de los formularios y procesarlos para ingresarlos a una base de datos. El procesamiento consiste en verificar (i) que la numeración del formulario no está en la base de datos para evitar duplicar formularios cargados, (ii) el formato de los campos de fechas, (iii) el flujo del formulario, (iv) la obligatoriedad de ciertos campos como el informante, la dependencia, el objetivo de la investigación, etc. (v) y la sintaxis de los campos para listar y para describir variables.

El proceso de carga se realiza con base en la numeración para mantener un orden que permita verificar con mayor facilidad el correcto ingreso en la base de datos, aunque puede cargarse de manera distribuida, por diferentes usuarios y en cualquier orden. Una vez los formularios han sido cargados pueden consultarse a través de la herramienta, y editarse si el usuario tiene los permisos para hacerlo.

En la figura 1 se muestra el diagrama del diseño de las tablas de F1e donde se alojan los formularios procesados y en la figura 2 los de F2e.



Figura 1: Diseño base de datos para F1e

# diagram BD F2e.png

Figura 2: Diseño base de datos para F2e

# Procesamiento

El procesamiento tiene como objetivo transformar los datos recopilados a través del diligenciamiento de los formularios F1e y F2e en información útil. Con los formularios se pueden identificar dos grandes transformaciones a hacerse con la base de datos recolectada:

* Procesar los datos de la encuesta para caracterizar las operaciones estadísticas con base en los módulos que describen el proceso estadístico agregando las respuestas a cada pregunta para toda la entidad y/o por dependencia.
* Construir un modelo del proceso de producción de información que relaciona los usuarios, los productores, las operaciones estadísticas, la estructura de la Alcaldía, el modelo de operación por procesos, y los productos de información, con miras a entender la estructura del Sistema Estadístico Municipal.

Ambas objetivos suponen un trabajo de procesamiento y comparten etapas. El proceso de agregación de la encuesta está definido en la documentación correspondiente al diseño estadístico. La construcción del modelo se encuentra en fase de diseño. En este documento se describe los procesos comunes que permiten la agregación y la construcción del modelo del sistema estadístico municipal.

## Preparación de los datos

Dado que el proceso de encuesta se propone explorar un universo de entidades en el rol de usuarios o productores e identificar dependencias y los grupos de trabajo, los campos de identificación son abiertos para ser ingresados como texto. Esto supone actualizar los campos con valores estandarizados e.g los nombres de las dependencias, el uso de clasificaciones y nomenclaturas que existan en la entidad, o extraer de los datos un conjunto de valores únicos de identificación, o crear nuevos campos con categorías estandarizadas que capturen características relevantes de los datos. Las variables que se someten a este procesamiento representan las unidades de observación y las unidades de análisis escogidas para el estudio; este procedimiento permite su agregación y la generación de consultas.

Otro aspecto importante de este pre-procesamiento es el uso de la sintaxis diseñada para el diligenciamiento y que permite automatizar la tabulación de listados ingresados en un campo del formulario separando los elementos con el carácter punto y coma ( ; ).

##

## Análisis exploratorio de los datos

Este proceso se basa en la generación de gráficos y cálculo de estadísticas que permiten familiarizarse con la naturaleza de los datos a analizar, estudiando las principales características de la distribución de las variables, valores atípicos, saltos o discontinuidades, concentraciones de valores. Con esto se busca poner de manifiesto las relaciones más evidentes que pudieran existir entre las variables y definir el mejor modelo o la forma de representar y visualizar la población de la cual proceden los datos.

El primer paso de este análisis consiste en el cálculo de los conteos de las respuestas a las preguntas cerradas por dependencia, y para toda la entidad. Esta agregación se hace usando el lenguaje de consulta estructurado o SQL (por sus siglas en inglés Structured Query Language) a la base de datos (se anexa el paquete de scripts de las consultas) y los resultados se exportan a un archivo separado por comas (csv). Los archivos se importan en una hoja de cálculo de Google que funciona bien como plataforma libre de procesamiento y visualización de datos.

Para la creación de gráficos las hojas de cálculo de Google permiten hacer una gran variedad de ellos[[2]](#footnote-1). Para la sistematización de procesos de preparación de los datos y automatización de tareas repetitivas se cuenta con Google Apps Script[[3]](#footnote-2), un lenguaje para ejecutar secuencias de comandos basado en JavaScript. Adicionalmente se usan las Fusion Tables de Google[[4]](#footnote-3) para hacer gráficos de red usados para mostrar las relaciones entre variables, en especial la concerniente a los flujos de información entre usuarios y productores.

El resultado de estos procesos ajustan el diseño de la visualización de datos y los cuadros de salida[[5]](#footnote-4), la identificación y corrección de datos anómalos, identificación de campos no respondidos, y finalmente se obtiene una base de datos actualizada y corregida.

## Producción de tablas y agregados estadísticos

La producción de los agregados y tablas definitivas ocurre después de la verificación por parte del equipo de recolección, del análisis de la coherencia de los resultados obtenidos y las mejoras en la presentación de los resultados con base en las recomendaciones realizadas .

1. En el documento “Guia para diligenciamiento de los formularios F1e y F2e en Excel V1.1” se encuentra la descripción de los formularios y la sintaxis para su diligenciamiento y la asignación de los nombres de los archivos. [↑](#footnote-ref-0)
2. Para mayor información ver la documentación oficial de Google. https://support.google.com/docs/topic/30240?hl=es&ref\_topic=1361474 [↑](#footnote-ref-1)
3. Google Apps Script https://developers.google.com/apps-script/overview [↑](#footnote-ref-2)
4. Fusion Tables https://support.google.com/fusiontables/answer/2571232 [↑](#footnote-ref-3)
5. Se anexan unos documentos en borrador llamados: Reporte de Oferta y Demanda, Reporte de Usuarios y Productores, Reporte del proceso de recolección, y que muestran algunos aspectos del diseño de los cuadros de salida y los reportes. [↑](#footnote-ref-4)