

INFORME ELECCIONES & | CONTRATOS

TRANSPARENCIA POR COLOMBIA

2018 - 2019



CONSEJO RECTOR

Margarita Garrido Otoyá
Rosa Inés Ospina Robledo
Carlos Angulo Galvis
José Alejandro Cortés Osorio
Andrés Echavarría Olano
Guillermo Carvajalino Sánchez

JUNTA DIRECTIVA

Rosa Inés Ospina Robledo
Eulalia Arboleda de Montes
Janet Bonilla Torres
Daniel Ricardo Uribe Parra
Bernardo Rodríguez Osa
Eduardo Wills Herrera
Néstor Ricardo Rodríguez Ardila

DIRECCIÓN EJECUTIVA

Gerardo Andrés Hernández Montes

DIRECCIÓN DE GESTIÓN DE CONOCIMIENTO

Marcela Restrepo Hung

AUTORES/ EQUIPO INVESTIGADOR

Sandra Ximena Martínez, Líder Sistema Político
Sergio Nicolás Rocha Camargo, Investigador
Camilo Peña Galeano, Investigador
Juan Sebastián Botero Romero, Ingeniero de Datos,
Leidy Cáceres Casilimas, Consultora

DISEÑO Y DIAGRAMACIÓN

Cristian Mauricio Ruiz Parra

CON EL APOYO FINANCIERO DE:

Open Society Foundation – OSF
Luminate

El contenido de este informe es responsabilidad exclusiva de la Corporación Transparencia por Colombia y no refleja la opinión de sus financiadores.

CIENCIA DE DATOS

Datasketch

Somos el Capítulo Nacional de Transparencia Internacional -TI-, la Organización de la Sociedad Civil líder en el mundo en la lucha contra la corrupción que trabaja en más de 100 países.

© Corporación Transparencia por Colombia

CORPORACIÓN TRANSPARENCIA POR COLOMBIA



Carrera 45A # 93-61, Barrio La Castellana
PBX: 610 0822
transparencia@transparenciacolombia.org.co
www.transparenciacolombia.org.co
Bogotá, Colombia.

08

ANEXOS

ANEXO 1. METODOLOGÍA



DESCRIPCIÓN DE LOS LIBROS

CAMPAÑAS:

Libro de datos donde se compilan y registran las tablas de Ingresos de Campañas electorales territoriales 2015, congreso 2018, presidenciales 2018 (1 y 2 vuelta), con los registros de las contribuciones enmarcadas dentro de los códigos 101 y 102*, proviene de las hojas de datos exportadas de la base de datos de Cuentas Claras.

FUNCIONAMIENTO:

Libro de datos donde se compilan y registran las tablas de aportes a Funcionamiento de partidos para los años 2016, 2017 y 2018, con los registros de las contribuciones enmarcadas dentro de los códigos 101 y 102*, proviene de las hojas de datos exportadas de la base de datos de Cuentas Claras.

CONTRATISTAS FINANCIADORES A CAMPAÑAS:

Libro de datos donde se compilan y registran las tablas de Ingresos de Campañas electorales territoriales 2015, congreso 2018, presidenciales 2018 (1 y 2 vuelta), con los registros de las contribuciones enmarcadas dentro de los códigos 101 y 102*, correspondiente a los financiadores de campaña que son a su vez contratistas del estado; proviene de las hojas de datos exportadas de la base de datos de Cuentas Claras y de las hojas de datos exportadas de los datasets de Colombia Compra Eficiente para SECOP I y SECOP II.

CONTRATOS DE FINANCIADORES A CAMPAÑAS:

Libro de datos donde se compilan y registran las tablas de contratos con el estado relacionadas en los sistemas SECOP I y SECOP II, correspondientes a los contratistas y representantes

legales relacionados como financiadores a Campañas electorales territoriales 2015, congreso 2018, presidenciales 2018 (1 y 2 vuelta), con los registros de las contribuciones enmarcadas dentro de los códigos 101 y 102*; proviene de las hojas de datos exportadas de los datasets de Colombia Compra Eficiente para SECOP I y SECOP II y las de hojas de datos exportadas de la base de datos de Cuentas Claras.

CONTRATOS DE FINANCIADORES DE PARTIDOS:

Libro de datos donde se compilan y registran las tablas de contratos con el estado relacionadas en los sistemas SECOP I y SECOP II, correspondientes a los contratistas y representantes legales relacionados como financiadores de Funcionamiento de partidos para los años 2016, 2017 y 2018, con los registros de las contribuciones enmarcadas dentro de los códigos 101 y 102*; proviene de las hojas de datos exportadas de los datasets de Colombia Compra Eficiente para SECOP I y SECOP II y las de hojas de datos exportadas de la base de datos de Cuentas Claras.

CONTRATISTAS FINANCIADORES DE PARTIDOS:

Libro de datos donde se compilan y registran las tablas de aportes a Funcionamiento de partidos para los años 2016, 2017 y 2018, con

los registros de las contribuciones enmarcadas dentro de los códigos 101 y 102*, correspondiente a los financiadores de partidos que son a su vez contratistas del estado; proviene de las hojas de datos exportadas de la base de datos de Cuentas Claras y de las hojas de datos exportadas de los datasets de Colombia Compra Eficiente para SECOP I y SECOP II.

CONTRATOS DE FINANCIADORES DE PARTIDOS:

Libro de datos donde se compilan y registran las tablas de contratos con el estado relacionadas en los sistemas SECOP I y SECOP II, correspondientes a los contratistas y representantes legales relacionados como financiadores de Funcionamiento de partidos para los años 2016, 2017 y 2018, con los registros de las contribuciones enmarcadas dentro de los códigos 101 y 102*; proviene de las hojas de datos exportadas de los datasets de Colombia Compra Eficiente para SECOP I y SECOP II y las de hojas de datos exportadas de la base de datos de Cuentas Claras.



PROCESO DE LIMPIEZA BASES DE DATOS CUENTAS CLARAS

Las hojas fueron recibidas originalmente como archivos planos CSV - UTF8 y reconvertidas a hojas de cálculo en formato XLSX, se filtraron los datos correspondientes a los ingresos registrados dentro de los *códigos 101 (Créditos o aportes que provengan del patrimonio de los candidatos, de sus cónyuges o de sus compañeros permanentes o de sus parientes) y 102 (Contribuciones, donaciones y créditos, en dinero o especie, que realicen los particulares); les aplicó un proceso de verificación de su estructura, normalizándose sus celdas de acuerdo a los tipos de datos, convirtiendo las celdas que contienen valores representados en pesos en celdas con formato contable, moneda peso y sin decimales, una vez realizado este proceso, se recompilaron las hojas en un sólo libro (para cada universo de datos) consolidándose los intervalos totales de datos en tablas, con el fin de facilitar el análisis de los datos y usar referencias estructuradas en vez de referencias de celda, todas las tablas han sido renombradas con referencias a su contenido.

PROCESO DE VERIFICACIÓN Y NORMALIZACIÓN BASES DE DATOS CUENTAS CLARAS

Se tomaron los números de identificación de las personas naturales y jurídicas aportantes a campaña y funcionamiento de partidos y se verificaron en su integridad como tipo de dato número, se limpiaron con fórmulas para eliminar espacios y caracteres no imprimibles y se retiraron caracteres convencionales como puntos, comas, etc, se hizo separación de columnas para retirar los divisores de identificación NIT y se eliminaron para esta tablas las identificaciones consideradas como inválidas, estas son consistentes con números en unidades o en decenas, los números registrados como repetición de 3 o más cifras en toda la extensión del número y las identificaciones registradas con 0, así mismo, se retiraron los ingresos registrados por personas sin números de identificación.

Estos números de identificación fueron verificados en forma masiva en el caso de personas naturales a través del sistema de verificación de antecedentes de

METODOLOGÍA DE ANÁLISIS

- El periodo de tiempo a analizar la contratación se da con aquellos contratos efectuados entre el 1 de enero de 2015 hasta el 30 de septiembre de 2019.
- Las campañas electorales a analizar son Territoriales 2015, Congreso 2018, Presidenciales 2018 (1ª y 2ª vuelta)
- Se determina como base de valores de contratación el campo de Valor Total más adicione
- Se descartan procesos de contratación registrados más de 1 vez asociados al mismo contratista, objeto, identificador y monto contratado.
- En el análisis se hace referencia explícita a aquellas personas naturales y/o jurídicas, financiadores de campañas, que aparezcan relacionados bien sea como contratistas y/o representantes legales de empresas contratistas.
- Se descartan procesos de contratación que se encuentren en estado Cancelado.
- Se agrupan los tipos de contratación en 7 categorías (Concurso De Méritos, Contratación directa, Licitación, Mínima cuantía, Otros, Régimen especial, Selección abreviada)
- Se agrupa el orden de la entidad contratante en 4 categorías ((Nacional, Departamental, Municipal-Distrital y No Definido).
- Al no existir una referencia definida de los territorios donde se ejecutan los contratos, se implementa una columna con datos de Departamento de ejecución del contrato mediante asociaciones de datos de los municipios registrados como Municipios de Ejecución.
- En las hojas de datos de Aportantes a campaña se establecen diferenciaciones entre candidatos aportantes y aportantes independientes.
- Se incorporan las bases de datos SECOP I y SECOP II, asociando sus columnas de datos con la similitud que corresponde entre ellas.



la Procuraduría, el sistema de consulta de estado de RUT de la DIAN, el sistema ADRES de registro de seguridad social y en el caso de personas jurídicas en los sistemas de consulta de RUT de la DIAN y de consulta en RUES, las consultas arrojan los nombres de las personas naturales y jurídicas tal y como están registradas en las bases de datos públicas gubernamentales, esta verificación fue consistente con los nombres y apellidos registrados en cada tupla o registro de aporte a campaña, por lo que las personas (naturales y jurídicas) que han hecho 2 o más aportes a una o más campañas se encuentran registradas con idéntico nombre.

De la misma manera, se verificó la consistencia y similaridad de los nombres de forma automatizada, para encontrar de forma relacional los nombres y apellidos en cualquier orden y verificar que el documento de identificación correspondiese a la persona que dice estar registrada, generándose una lista alterna de identificaciones inexistentes o no correspondientes, que se verificaron de nuevo para encontrar los casos de errores de digitación en el documento y corregir el mismo. El número de documento corregido y el nombre verificado y normalizado se anexaron en 2 nuevas columnas, dejando los datos originales registrados para verificaciones más profundas.

En el caso de la aparición de personas cuya información no se encuentra disponible en Procuraduría y ADRES, de forma ordinaria, se hizo verificación del número de identificación cuando este consiste en 6 cifras, que por lo regular corresponde a cédulas de extranjería, identificando los ciudadanos poseedores de tal documento y diferenciándolos, puesto que las cédulas nacionales que corresponden a estos números por lo general corresponde a ciudadanos ya fallecidos. Los ciudadanos cuyas cédulas de ciudadanía no corresponden a documentos de extranjería, fueron verificados a través del sistema de consulta de Censo Electoral y Puestos de Votación de la Registraduría Nacional, verificando que la cédula esté activa o ya dada de baja por muerte del titular. En otros casos, número de cédula del orden de 10 cifras, pueden no aparecer como registrados en bases de datos de antecedentes SIRI por corresponder a personas que han adquirido su mayoría de edad muy recientemente, estas personas fueron verificadas en la base de datos ADRES que registra su nombre asociado a una EPS, en el caso de los varones se identificaron varios registros a través del sistema de consulta de Situación Militar que ofrece la dirección Nacional de Reclutamiento de las FFMM de Colombia.



PROCESO DE LIMPIEZA BASES DE DATOS SECOP UNO Y DOS

Se abordaron los registros de contratos públicos de las plataformas de contratación pública SECOP I y SECOP II, con un enfoque de análisis de información pública. La información se tomó del portal <http://datos.gov.co> donde se cuenta la información con posibilidad de descarga masiva. Los vínculos pueden ser consultados directamente en la página de Colombia Compra Eficiente aquí: <https://www.colombiacompra.gov.co/transparencia/conjuntos-de-datos-abiertos>

El SECOP está conformado por una plataforma de publicación de información (SECOP I) que cuenta con más de 8 millones de registros de procesos de contratación y dos plataformas transaccionales, el SECOP II que cuenta con más 350.000 publicados a la fecha y la Tienda Virtual del Estado Colombiano. Este reporte utiliza la información de los sistemas SECOP I y SECOP II.

Para explicar las conexiones entre la contratación pública y el financiamiento de campañas políticas fue necesario descargar la información masiva e incorporar diferentes rutinas de limpieza de información en los más de 8 millones de registros.

Limpieza y preparación de información

Esta base de datos cuenta con el mayor número de registros. Estos registros son introducidos de forma manual por diferentes encargados en cada una de las entidades públicas. Debido a este procedimiento de captura se presentan bastantes errores en la información reportada que dificultan el trabajo de veedores públicos o investigadores. Uno de los problemas principales cuando se analizan datos, sobre todo aquellos que proveniente de múltiples fuentes, es la cantidad de contenidos faltantes. Muchos de los campos más importantes están vacíos o mal diligenciados, campos vacíos tienen diferentes cadenas de caracteres para indicar información faltante.

Con el fin de preparar la información lo mejor posible para los cruces de información primero se crea un diccionario de datos. Este diccionario permite la estandarización de nombres de variables como encabezados para facilitar la lectura automatizada por máquinas. Una vez están normalizados los encabezados se pasa a una rutina de limpieza de información que incorpora posibles correcciones para casos recurrentes que cumplen alguno de los siguientes criterios:

- Cédulas con múltiples formatos, por ejemplo con puntos o sin puntos.
- Personas con diferentes nombres y apellidos que tienen el mismo número de cédula.
- La misma persona con su nombre escrito con tildes, sin tildes y cualquier combinación de mayúsculas y minúsculas.
- Nombres de representantes legales que en realidad son nombres de empresas.
- Cédulas o NITs escritos completamente con ceros (00000000) o unos (111111111)
- Cédulas o NITs escritos con los nombres en lugar de los números de identificación de las personas o las empresas.
- Mismos contratistas con diferentes nombres por problemas de digitación.
- Falta de estandarización de siglas de empresas.
- Información escrita en la columna equivocada. Por ejemplo, nombres en campos de cédulas o teléfonos en direcciones.
- Municipios de ejecución que no tienen código DANE en la base de datos de SECOP.

Para los puntos mencionados anteriormente se implementaron diversas rutinas de limpieza y estandarización de información, incluyendo:

- Clasificación de identificaciones para revisión posterior sobre si un número corresponde a una cédula de persona natural o a un NIT de persona jurídica.
- Remoción de caracteres especiales, conversión a formato de cédula o NIT separado del dígito de verificación.
- Todas las validaciones que tienen que ver con el número de cédula se realizan con respecto al campo de cédula de contratista cuando es persona natural, así como con el campo de cédula del representante legal cuando el contratista es una empresa.
- Validación de fechas: en diversas oportunidades se han encontrado en los registros de bases de datos información de fechas que no existen, por ejemplo, 31 de noviembre.
- Estandarización de códigos DANE para regiones y municipios. Este ejercicio se realiza a dos niveles un primer nivel sobre las variables de departamentos y municipios, y dos en los casos donde múltiples municipios están en una misma celda se crea un nuevo campo de código DANE que conserva el orden de los municipios.
- Opcional se puede incorporar el cruce de información con nombres comunes de personas en español para identificar el género de los contratistas. Futuras versiones de la aplicación podrían considerar la incorporación de esta información.



Adicionalmente se revisó en el transcurso de la investigación aquellos casos que representaban alguna irregularidad evidente al analizar la información o explorar la herramienta de visualización desarrollada. Estos datos identificados como fuera de serie permiten encontrar registros de información irregulares, el principal caso de este tipo son aquellos en los que la información del monto de los contratos es ingresada de manera errónea, haciendo que la información de los totales de los contratos cambie considerablemente en casos donde se incorporan, por ejemplo, los contratos de pasan de millones a billones debido a que se ingresaron más ceros de los necesarios. Estos casos requieren un tratamiento particular, ya que al ser datos oficiales como equipo de investigación no podemos entrar a editarlos, en estos casos se recurre a hacer una marcación de estos registros en las bases de datos para que las entidades responsables conozcan los casos para su eventual resolución.

En el caso de SECOP II se realizó un tratamiento similar para las variables relevantes. En un primer acercamiento se realizó la unificación de SECOP DOS con SECOP UNO, para lo cual fue necesario renombrar la tabla del SECOP DOS de acuerdo a los datos del diccionario asignado al SECOP UNO. Al realizar la consolidación de información de los dos sistemas de contratación, el número de variables no necesariamente coincide. Las variables comunes se mantienen y se agregan aquellas nuevas que fueron relevantes para la aplicación que estaban en SECOP II pero que no se encontraban originalmente en SECOP I. En particular para el cruce se tuvo en cuenta la estandarización de aquellas variables que coincidían pero que no estaban en el mismo formato como por ejemplo: "Fecha de Inicio del Contrato" (variable SECOP DOS) que tiene su símil con "Anno firma del contrato" (variable SECOP UNO), en este caso se separó la fecha del segundo para adicionar la variable comparable en el primero. En general se verificó que las variables de identificación no tuvieran problemas evidentes de formato al unirse. Finalmente se realizó el filtro de todos los contratos públicos disponibles y normalizados de manera que tuviéramos únicamente aquellos contratos pertenecientes a financiadores de campañas o de funcionamiento de partidos, según información de Cuentas Claras. Este filtro se realizó con respecto a las identificaciones de contratistas (cédulas para personas naturales y NITs para empresas) así como las cédulas de representantes legales de empresas.

METODOLOGÍA DE ANÁLISIS

- El periodo de tiempo a analizar la contratación se da con aquellos contratos efectuados entre el 1 de enero de 2015 hasta el 30 de septiembre de 2019.
- Las campañas electorales a analizar son Territoriales 2015, Congreso 2018, Presidenciales 2018 (1ª y 2ª vuelta)
- Se determina como base de valores de contratación el campo de Valor Total más adicione
- Se descartan procesos de contratación registrados más de 1 vez asociados al mismo contratista, objeto, identificador y monto contratado.
- En el análisis se hace referencia explícita a aquellas personas naturales y/o jurídicas, financiadores de campañas, que aparezcan relacionados bien sea como contratistas y/o representantes legales de empresas contratistas.
- Se descartan procesos de contratación que se encuentren en estado Cancelado.
- Se agrupan los tipos de contratación en 7 categorías (Concurso De Méritos, Contratación directa, Licitación, Mínima cuantía, Otros, Régimen especial, Selección abreviada)
- Se agrupa el orden de la entidad contratante en 4 categorías ((Nacional, Departamental, Municipal-Distrital y No Definido).
- Al no existir una referencia definida de los territorios donde se ejecutan los contratos, se implementa una columna con datos de Departamento de ejecución del contrato mediante asociaciones de datos de los municipios registrados como Municipios de Ejecución.
- En las hojas de datos de Aportantes a campaña se establecen diferenciaciones entre candidatos aportantes y aportantes independientes.
- Se incorporan las bases de datos SECOP I y SECOP II, asociando sus columnas de datos con la similitud que corresponde entre ellas.

TABLA 9. CONTENIDO DE DATOS CAMPAÑAS

FECHA DE ÚLTIMO CORTE REGISTRADO			
31/05/2019			
PERIODO DE ANÁLISIS			
01/10/2015			
12/09/2019			
TABLA	DATOS	FILAS	COLUMNAS
CAMPAÑAS ELECTORALES	2.639.345	143.747	24

TABLA 10. CONTENIDO DE DAOS FUNCIONAMIENTO

FECHA DE ÚLTIMO CORTE REGISTRADO			
19/03/2019			
PERIODO DE ANÁLISIS			
01/01/2016			
29/04/2019			
TABLA	DATOS	FILAS	COLUMNAS
FUNCIONAMIENTO	109.710	12.245	12

TABLA 11. CCONTENIDO DE DATOS FINANCIADORES A CAMPAÑAS CONTRATISTAS:

FECHA DE ÚLTIMO CORTE REGISTRADO			
31/05/2019			
PERIODO DE ANÁLISIS			
01/10/2015			
12/09/2019			
TABLA	DATOS	FILAS	COLUMNAS
APORTANTES-CAMPANA	783.725	34.075	23

TABLA 12. CONTENIDO DE DATOS CONTRATOS DE FINANCIADORES A CAMPAÑAS

FECHA DE ÚLTIMO CORTE REGISTRADO			
30/09/2019			
PERIODO DE ANÁLISIS			
01/01/2015			
30/09/2019			
TABLA	DATOS	FILAS	COLUMNAS
CONTRATOS-CAMPANA	13.152.755	170.815	77

TABLA 13. CONTENIDO DE DATOS FINANCIADORES DE PARTIDOS CONTRATISTAS

FECHA DE ÚLTIMO CORTE REGISTRADO			
31/05/2019			
PERIODO DE ANÁLISIS			
01/01/2016			
12/09/2019			
TABLA	DATOS	FILAS	COLUMNAS
APORTANTES-FUNCIONAMIENTO	13.090	935	14

TABLA 14. CCONTENIDO DE DATOS CONTRATOS DE FINANCIADORES DE PARTIDOS

FECHA DE ÚLTIMO CORTE REGISTRADO			
30/09/2019			
PERIODO DE ANÁLISIS			
01/01/2015			
30/09/2019			
TABLA	DATOS	FILAS	COLUMNAS
CONTRATOS-FUNCIONAMIENTO	670.683	8.755	77

