

Departamento Administrativo  
Nacional de Estadística



Diseño  
DSO

DIRECCION DE CENSOS Y DEMOGRAFIA  
DCD

**METODOLOGÍA GENERAL PARA LA  
CONSTRUCCIÓN Y ACTUALIZACIÓN DEL  
REGISTRO ESTADÍSTICO BASE DE  
POBLACIÓN – REBP 2018**

Diciembre 2019

**TABLA DE CONTENIDO**

<b>LISTA DE TABLAS</b> .....	<b>4</b>
<b>LISTADO DE FIGURAS</b> .....	<b>4</b>
<b>PRESENTACIÓN</b> .....	<b>5</b>
<b>INTRODUCCIÓN</b> .....	<b>6</b>
<b>1. ANTECEDENTES</b> .....	<b>7</b>
<b>2. MARCO DE REFERENCIA</b> .....	<b>8</b>
<b>2.1. OBJETIVOS</b> .....	<b>8</b>
2.1.1 OBJETIVO GENERAL .....	8
2.1.2 OBJETIVOS ESPECÍFICOS .....	8
<b>3. ESPECIFICACIÓN DE NECESIDADES DE INFORMACIÓN</b> .....	<b>10</b>
<b>4. CONSTRUCCIÓN</b> .....	<b>13</b>
<b>5. PROCESAMIENTO</b> .....	<b>14</b>
<b>5.1. DEFINICIÓN DE VARIABLES</b> .....	<b>14</b>
5.1.1. VARIABLES DE IDENTIFICACIÓN REBP .....	14
5.1.2. VARIABLES DE UBICACIÓN .....	15
5.1.3. VARIABLES DEMOGRÁFICAS .....	15
5.1.4. DERIVACIÓN DE NUEVAS VARIABLES .....	15
<b>5.2 CODIFICACIÓN, HOMOLOGACIÓN Y VALIDACIÓN PARA CADA REGISTRO ADMINISTRATIVO FUENTE</b> .....	<b>18</b>
5.2.1 DEPURACIÓN E IMPUTACIÓN DE LA VARIABLE SEXO .....	20
5.2.2 DEFINICIÓN DE DUPLICADOS POR ANÁLISIS DETERMINÍSTICO .....	20
<b>5.3 EMPAREJAMIENTO, INTEGRACIÓN Y MEJORAMIENTO DE VARIABLES DEL REBP</b> .....	<b>20</b>
<b>6. ANÁLISIS Y VALIDACIÓN DE RESULTADOS</b> .....	<b>23</b>
<b>7. PRODUCTOS DE DIFUSIÓN</b> .....	<b>24</b>
<b>8. GLOSARIO</b> .....	<b>25</b>
<b>9. BIBLIOGRAFÍA</b> .....	<b>28</b>
<b>10. ANEXOS</b> .....	<b>29</b>

**LISTA DE TABLAS**

<b>Tabla 1.</b> Listado de registros administrativos para la conformación del REBP.....	<b>9</b>
<b>Tabla 2.</b> Ejemplo de conversión de carácter a fonético.....	<b>15</b>

**LISTADO DE FIGURAS**

<b>Figura 1.</b> Los cuatro registros Base.....	<b>11</b>
<b>Figura 2.</b> Proceso de construcción de un Registro Estadístico Base de Población.....	<b>11</b>
<b>Figura 3.</b> Codificación y Reglas de validación.....	<b>17</b>
<b>Figura 4.</b> Proceso de Emparejamiento e Integración de Registros.....	<b>19</b>

## PRESENTACIÓN

El Departamento Administrativo Nacional de Estadística (DANE), en el marco del proyecto de Planificación y Armonización Estadística, trabaja por el fortalecimiento y la consolidación del Sistema Estadístico Nacional.

En este contexto, la función del DANE es coordinar la producción de estadísticas estratégicas; la generación, adaptación, adopción y difusión de estándares; la consolidación y armonización de la información estadística y la articulación de instrumentos, actores, iniciativas y productos. Estas acciones tienen como fin mejorar la calidad de la información estadística estratégica, su disponibilidad, oportunidad y accesibilidad para responder a la gran demanda que se tiene de ella.

Consciente de la necesidad de brindar a los usuarios mejores productos, el DANE generó una guía para la presentación de metodologías que contribuye en la visualización y entendimiento del proceso estadístico. Con este instrumento, el DANE pone a disposición de los usuarios especializados y el público en general, los documentos metodológicos de sus operaciones e investigaciones estadísticas, que permitan de manera completa y de fácil lectura las principales características técnicas de los procesos y subprocesos de cada investigación, lo que permite su análisis, control, posibilidad de ser replicado y evaluación.

Esta serie de documentos favorecen la transparencia, la confianza y la credibilidad de la calidad técnica de la institución para un mejor entendimiento, comprensión y aprovechamiento de la información estadística. Tal información es producida bajo los principios de coherencia, comparabilidad, integralidad y calidad de las estadísticas.

## INTRODUCCIÓN

El Departamento Administrativo Nacional de Estadística - DANE – tiene como misión, garantizar la producción y difusión de información estadística estratégica de calidad para la toma de decisiones y la investigación en Colombia.

A medida que los países alrededor del mundo se acercan a la adopción de los Objetivos de Desarrollo Sostenible, se ha hecho un llamado a los gobiernos, sociedad civil, ciudadanos y al sector privado a unirse a una revolución de datos. Esta revolución trae consigo grandes retos, posibilidades y nuevas demandas de información. Como respuestas a estas demandas, actualmente se dispone de un gran arsenal de fuentes que producen datos muy detallados producidos en tiempo real, en una variedad de formatos, procedentes de fuentes diversas, y que recogen información valiosa sobre innumerables aspectos de la vida de las personas, el funcionamiento de las organizaciones, el estado del medio ambiente en que vivimos, y todas sus interacciones. En medio de este proceso, surge la necesidad urgente de aprovechar las oportunidades que trae esta revolución para mejorar el diseño y ejecución de políticas públicas y la rendición de cuentas, con el objetivo último de mejorar la vida de los ciudadanos de todo el mundo.

De otro lado, los gobiernos requieren de información oportuna y de buena calidad que les permita tomar decisiones eficaces basadas en evidencia. Es aquí donde los registros administrativos emergen como un complemento importante de los censos y las encuestas tradicionales. Los registros administrativos permitirían una mayor eficiencia en el proceso de recolección de datos y una cobertura temática y nivel de detalle de datos más amplios. Alcanzar este potencial requiere el fortalecimiento de la capacidad institucional de los Sistemas Estadísticos Nacionales para captar información útil de una mejor manera, mediante el aprovechamiento de las innovaciones tecnológicas y la estandarización de conceptos en los organismos gubernamentales. En esta tarea, los gobiernos hacen frente a limitantes legales, técnicas y tecnológicas. Por ejemplo, mientras los gobiernos buscan facilitar el intercambio de información a través de modificaciones a sus marcos legales y mejor coordinación, deben también proteger la privacidad de los ciudadanos y las empresas. Con distintos grados de avance, los gobiernos de diferentes países buscan un equilibrio entre estas variables para así mejorar la calidad de la información estadística derivada de los registros administrativos.

En esta revolución de datos, son las Oficinas Nacionales de Estadística los llamados a ser agentes catalizadores en la realización del potencial de los registros administrativos, no sólo debido a su rol natural de entes coordinadores y reguladores de la calidad de las estadísticas oficiales, sino también como mensajeros principales de la importancia de los datos y las estadísticas.

Bajo este lineamiento y dando respuesta a la necesidad de fortalecer el uso de registros administrativos con fines estadísticos, se ha propuesto una metodología para conformar un registro estadístico base de población, documentando todas las etapas necesarias para transformar un registro administrativo de población en registro estadístico (evaluación de la cobertura, revisión de las variables incluidas, calidad de la información y frecuencia de disponibilidad de la información) y la forma en que este fue conformado. La metodología podrá ser aplicada de forma general a cualquier territorio del país.

	<b>Metodología general para la construcción y actualización del Registro Estadístico Base de Población - REBP</b>	CÓDIGO: VERSIÓN: 01
PROCESO: Misional	SUBPROCESO: Diseño	

## 1. ANTECEDENTES

En la última década se han generado a nivel global diferentes iniciativas para el mejoramiento en la producción de información estadística. La primera de ellas, corresponde a la promovida por la Organización de las Naciones Unidas (ONU), la Comisión Europea, la Organización para la Cooperación y el Desarrollo Económico (OCDE), el Fondo Monetario Internacional (FMI) y el Banco Mundial con la creación del Consorcio de Estadísticas para el Desarrollo en el Siglo XXI (Partnership in Statistics for Development in the 21<sup>st</sup> Century – PARIS 21) destinado a promover estrategias para el fortalecimiento y el mejor uso de las estadísticas.

Por otro lado, el uso de nuevas fuentes de información y la exploración de nuevos enfoques metodológicos para la producción de estadísticas basadas en registros administrativos, se han convertido en un tema prioritario a incorporar por parte de las Oficinas Nacionales de Estadística, específicamente en la implementación y el monitoreo de los Objetivos de Desarrollo Sostenible (ODS) hacia el desarrollo de la Agenda 2030.

Con relación a este último punto, los países nórdicos se destacan como pioneros en la utilización de registros administrativos, quienes desde los años 60, han encabezado esfuerzos en el uso de esta información en los procesos de producción estadística. Entre tanto, en otros países y en forma más reciente, se ha venido implementando el desarrollo de capacidades en sus profesionales e instituciones para el uso de este tipo de información, la cual se considera una fuente importante que puede ser incorporada en los procesos de producción de estadísticas oficiales.

Pero los registros administrativos en sí mismos, no son una fuente estadística propiamente dicha. Estos, corresponden al resultante de las necesidades de cada institución, creado con la finalidad de viabilizar la administración de los programas de gobierno o para fiscalizar el cumplimiento de obligaciones legales de la sociedad. Para su utilización, es imprescindible tener en cuenta su base conceptual y metodológica (CEPAL). Para su uso con fines estadísticos es preciso evaluar su base, clasificaciones, cobertura alcanzada, variables investigadas, calidad de las respuestas, procesamiento de los datos y frecuencia de disponibilidad de ellos.

Para el desarrollo de esta estrategia, es necesario elaborar metodologías y procedimientos para el uso eficiente de los registros administrativos en la producción de estadísticas oficiales, que incluyan la conversión de estos en registros estadísticos, con el propósito de incorporarlos a la infraestructura del Sistema Estadístico Nacional de cada país (Decisión 780 11 de marzo de 2013 – CAN).

En los últimos años, el DANE se ha incorporado a las iniciativas para el uso de los datos contenidos en los registros administrativos de las instituciones públicas y privadas, como insumo para la producción estadística convirtiéndolo en un potente instrumento para la medida y el análisis de la realidad en diversos ámbitos.

## 2. MARCO DE REFERENCIA

La Ley 1753 de 2015 en su artículo 160 creó el Sistema Estadístico Nacional para proporcionar a la sociedad y al Estado, estadísticas nacionales y territoriales de calidad con lenguajes y procedimientos comunes, siguiendo estándares estadísticos internacionales, que contribuyan a la transparencia, pertinencia, interoperabilidad, acceso, oportunidad y coherencia de las estadísticas producidas en el país. En este sentido, el SEN está integrado por todas las entidades que producen y difunden estadísticas o sean responsables de registros administrativos, por tanto, deben implementar los lineamientos, buenas prácticas, estándares y normas técnicas que el DANE defina soportado en referentes internacionales, para la producción y difusión de estadísticas oficiales y para el aprovechamiento estadístico de los registros administrativos.

Adicionalmente, el DANE ha proporcionado mediante el Decreto 1743 de 2016, los mecanismos para el fortalecimiento y aprovechamiento amplio e intensivo de los registros administrativos, como fuente para la producción de estadísticas oficiales y el mejoramiento de la calidad y coherencia de las mismas a partir de la coordinación de los miembros del SEN y organismos internacionales.

A nivel internacional, diferentes organizaciones están investigando la posibilidad de mejorar el uso de los registros administrativos, dado que proporcionan nuevas oportunidades en la producción de estadísticas, además de reducir la carga y los costos en la obtención de respuestas. La presente metodología ilustra el proceso realizado para construir un Registro Estadístico Base de Población-REPB, que permita conocer la dinámica de la población residente en Colombia, demandante de servicios a cargo de entidades públicas, exponiendo los procesos y procedimientos que tienen correspondencia con las etapas del GSBPM - Modelo Genérico del Proceso Estadístico.

### 2.1. Objetivos

#### 2.1.1 Objetivo general

Construir de forma periódica un insumo para determinar la enumeración y caracterización básica de la población, su estructura y distribución geográfica de forma longitudinal y trasversal, así como también para la construcción de supuestos de la dinámica poblacional que permitan actualizar las proyecciones de población en el período intercensal.

#### 2.1.2 Objetivos Específicos

1. Producir información insumo para la actualización demográfica, con una opción diferente de recolección de información, que permita caracterizar a la población dentro del período intercensal.

2. Promover el uso de fuentes no tradicionales, fortaleciendo la información de entidades de orden nacional y territorial, con el fin de evitar duplicidad de esfuerzos, mejorando la oportunidad en la entrega de resultados y reduciendo costos en la producción de estadísticas.
3. Consolidar el Registro Estadístico Base de Población, para avanzar hacia un sistema de información basado en registros.

	<b>Metodología general para la construcción y actualización del Registro Estadístico Base de Población - REBP</b>	CÓDIGO: VERSIÓN: 01
PROCESO: Misional	SUBPROCESO: Diseño	

### 3. ESPECIFICACIÓN DE NECESIDADES DE INFORMACIÓN

En esta etapa se identifican las necesidades de información estadística y se analizan los requerimientos que deben ser cubiertos, identificando los registros administrativos útiles al DANE para la conformación del registro estadístico base. Para esto se realiza una revisión previa de las posibles fuentes de datos administrativas (proveedores), viabilidad para su acceso y los mecanismos establecidos para acceder a estas fuentes de información.

Durante este proceso, se explora el inventario de registros administrativos, de la Dirección de Regulación, Planeación, Estandarización y Normalización DIRPEN, el cual incluye las características de los registros administrativos: su objetivo, las unidades de análisis y de observación, la cobertura; a cargo de las entidades públicas y privadas del orden nacional y territorial, que conforman el SEN y pueden ser aprovechados con fines estadísticos y en este caso, en la conformación del Registro Estadístico Base de Población- REBP

Así mismo, se revisan las variables y los conceptos que determinan las características principales de la población, básicas para la generación y análisis de estadísticas, teniendo en cuenta que la dirección de censos y demografía, es el principal usuario del REBP, contando con la experiencia adquirida durante el operativo censal del CNPV 2018, para el análisis de consistencias y producción de indicadores demográficos. En el cual se utilizaron los registros administrativos a cargo de la Registraduría Nacional del Estado Civil (RNEC), la Base Única de Afiliados (BDUA) y el Sistema de Identificación de Potenciales Beneficiarios de Programas Sociales (SISBEN), entre otros. Lo que posibilitó contar con un parámetro de referencia y validación con recolección tradicional de puerta a puerta y la propuesta de un Registro Estadístico conformado con registros administrativos, que permitirá la dinámica de analizar la poblacional durante el periodo intercensal.

Producto de esa etapa se identifican aquellos registros administrativos que aportan información demográfica, para la conformación del REBP como:

- Nombres y apellidos
- Tipo de documento
- Número de documento
- Fecha de nacimiento
- Sexo y una aproximación a la residencia habitual (departamento y municipio) de la población residente en Colombia, en su mayoría, bajo la responsabilidad de entidades públicas que prestan servicios de identificación, salud y educación.

Teniendo en cuenta que los registros administrativos priorizados son fuente de información para la producción de estadísticas vitales (nacimientos y defunciones) y la identificación de los miembros del hogar en las operaciones estadísticas dirigidas a hogares, incluyendo el censo de población y vivienda, también indagan las variables mencionadas anteriormente, la Gran Encuesta de Hogares y las estadísticas vitales, tanto en procesos y procedimientos, como en gestión y resultados esperados

fueron referencia para su reutilización en la conformación del REBP. Igualmente, se aprovecharon los convenios establecidos entre el DANE y las entidades a cargo de los registros administrativos para acceder a las bases de datos y para la visualización de resultados se aprovechó para el desarrollo del geovisor de la Dirección de Geoestadística.

Los registros administrativos requeridos para la conformación del REBP son:

**Tabla 1.** Listado de registros administrativos para la conformación del REBP

Registro Administrativo	Entidad de Origen	Convenio DANE - Forma de Acceso	Periodicidad de Entrega
RCN Registro Civil de Nacimiento ANI Archivo Nacional de Identificación	RNEC Registraduría Nacional del Estado Civil	A través de FTP	Sin establecer
BDUA Base Única de Afiliados del régimen Contributivo y Subsidiado BDEX Base Única de Afiliados del régimen de Excepción.	MSPS Ministerio de Salud y Protección Social	Convenio 996 de 2013 suscrito entre en el DANE y el Ministerio de Salud	Según solicitud, se espera periodicidad semestral
EPS Entidades Promotoras de Salud	EPS Entidades Promotoras de Salud	A través de Correo Electrónico o medio magnético	Según solicitud, se espera semestral
SISBEN Sistema de Selección de Beneficiarios Para Programas Sociales	DNP Departamento Nacional de Planeación	A través de FTP	Mensual desde abril de 2015
Estadísticas Vitales Nacimientos y Defunciones	MSPS Ministerio de Salud y Protección Social	Información Interna del DANE	Según solicitud
PILA Planilla Integrada de Liquidación de Aportes de Cotizantes y Pensionados	MSPS Ministerio de Salud y Protección Social	Convenio 996 de 2013 suscrito entre en el DANE y el Ministerio de Salud	Según solicitud
GEIH Gran Encuesta Integrada de Hogares	DANE	Información Interna del DANE	Según solicitud
ECV Encuesta de Calidad de Vida	DANE	Información Interna del DANE	Según solicitud
FFMM Fuerzas Militares CREMIL Caja de Retiro Militar	Ministerio de Defensa Nacional	A través de Medio Magnético	Según solicitud
Cedulas de Extranjería	Migración Colombia	A través de Medio Magnético	Según solicitud
SIMAT Sistema Integrado de Matriculas	MEN Ministerio de Educación Nacional	Convenio marco interinstitucional entre el Ministerio	Según solicitud, se espera

Registro Administrativo	Entidad de Origen	Convenio DANE - Forma de Acceso	Periodicidad de Entrega
SNIES Sistema Nacional de Información de la Educación Superior SIET Sistema de Información de Educación para el Trabajo		de Educación y el DANE.	periodicidad semestral
RLCPD Registro para la Localización y Caracterización de Personas con Discapacidad	MSPS Ministerio de Salud y Protección Social	Convenio 996 de 2013 suscrito entre en el DANE y el Ministerio de Salud	Según solicitud
SIGEP Sistema de Información y Gestión del Empleo Público	Departamento Administrativo de la Función Pública	A través de Medio Magnético	Según solicitud
RUV Registro Único De Víctimas	Unidad Para las Víctimas	A través de medio magnético	
RUT Registro Único Tributario	DIAN Dirección de Impuestos y Aduanas Nacionales	A través de medio magnético	
PAI Programa Ampliado de Inmunización Nacional	MSPS Ministerio de Salud y Protección Social	Convenio 996 de 2013 suscrito entre en el DANE y el Ministerio de Salud	Semestral
PAI Programa Ampliado de Inmunización Distrital	SDS Secretaria Distrital de Salud de Bogotá	A través de medio magnético	Semestral
RIPS Registros Individuales de Prestación de Servicios de Salud	MSPS Ministerio de Salud y Protección Social	Convenio 996 de 2013 suscrito entre en el DANE y el Ministerio de Salud	Semestral

El instrumento definido para la obtención del registro administrativo es una carta de solicitud a la entidad proveedora de información, proyectada por la Dirección Técnica y revisada y aprobada por la Subdirección del DANE, la cual contiene el listado de variables requeridas para el desarrollo del proyecto. En relación con los medios en los que pueden ser dispuestos los registros administrativos se encuentran: correo electrónico, medios magnéticos, USB y FTP entre otros. Se debe especificar además el formato en que será enviada la información. La frecuencia definida para la actualización de la información es anual.

Los datos solicitados a las entidades públicas y privadas, así como información derivada de los mismos, son confidenciales y su uso es exclusivo para fines estadísticos como lo establece la ley de Reserva Estadística (Ley 73 de 1993, artículo 4).

	<b>Metodología general para la construcción y actualización del Registro Estadístico Base de Población - REBP</b>	CÓDIGO: VERSIÓN: 01
PROCESO: Misional	SUBPROCESO: Diseño	

#### 4. CONSTRUCCIÓN

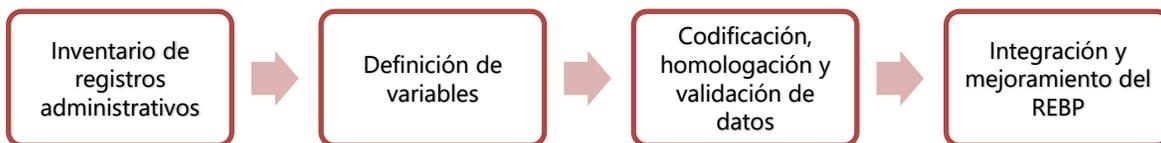
De acuerdo con la metodología para la generación de estadísticas a partir de registros administrativos de Wallgren & Wallgren (2012), el término registro base solo se debe utilizar para referirse a alguno de los cuatro registros que conforman el sistema basado en registros (ver figura 1)

Figura 1. Los cuatro registros base



La metodología de construcción de un Registro Estadístico Base de Población se establece en cuatro pasos, los cuales se detallarán en este documento.

Figura 2. Proceso de construcción de un Registro Estadístico Base de Población



Para la construcción del REBP, es fundamental contar con los registros administrativos generados por cada institución del orden nacional y territorial. Inicialmente, se realiza la consulta en el inventario de registros administrativos del DANE con el fin de identificar los datos de población que satisfacen la necesidad de información. Así mismo, se revisan los conceptos y variables requeridas. Una vez se cuenta con el acceso a las unidades y variables, estas se consolidan en una base de datos, donde se clasifica, codifica y estandariza la información.

	<b>Metodología general para la construcción y actualización del Registro Estadístico Base de Población - REBP</b>	CÓDIGO: VERSIÓN: 01
PROCESO: Misional	SUBPROCESO: Diseño	

## 5. PROCESAMIENTO

Una vez se reciben las bases de datos de los registros administrativos solicitados, las variables de interés se clasifican, codifican y estandarizan, de tal manera que permita la correlación de tablas de referencia, homologación de nombres de variables y recodifican categorías, se diseñan funciones para corrección de los caracteres especiales en los campos de identificación de las personas, para separación de nombres y apellidos cuando son reportados en un mismo campo, también se utiliza la expresión fonética y el conteo de las letras que contienen las variables de identificación.

Posteriormente se lleva a cabo el emparejamiento determinístico de las fuentes de información, que consiste en identificar registros que corresponden a la misma persona en dos o más archivos e integrarlos cuando estos coinciden, se realiza por medio de variables de vinculación que es un término que engloba tanto a variables de identificación como de ubicación. Un vínculo entre dos registros consta de una o más de estas llaves.

A continuación, se realiza una breve descripción de las variables que conforman el REBP. El Anexo 1 corresponde al diccionario de datos.

### 5.1. Definición de variables

#### 5.1.1. Variables de Identificación REBP

**Número de identificación estadístico:** Se definirán dos identificadores estadísticos, el primero se denomina ID\_TABLA el cual corresponde al consecutivo auto numérico asignado por el sistema para identificar cada registro administrativo ingresado al registro estadístico base de población. El segundo se denomina ID\_ESTADISTICO el cual corresponde al consecutivo auto numérico asignado por el sistema para identificar cada registro del proveedor de información.

Los identificadores estadísticos no necesariamente corresponden a una única persona puesto que en el análisis se estableció que la misma persona puede ser cargada varias veces en el sistema por error debido al cambio del documento o errores en los procesos de digitación del número de identificación.

**Número de documento:** (NUMERO\_IDENTIFICACION) Código de identificación asignado por la Registraduría Nacional del Estado Civil para los colombianos y asignado por el Ministerio del Exterior para los extranjeros, el cual permite la identificación inequívoca de las personas. Este número es la llave de cruce principal con los diferentes registros administrativos.

**Tipo de documento:** (ID\_TIPO\_IDENTIFICACION) Para los colombianos se manejan tres tipos de documentos de identidad (Registro Civil de 0 a 7 años, Tarjeta de Identidad de 7 a 18 años, y Cédula de Ciudadanía mayores de 18 años), para los extranjeros se maneja la Cédula de Extranjería, el Pasaporte y el Carnet Diplomático.

**Nombres y apellidos:** (PRIMER\_APELLIDO, SEGUNDO\_APELLIDO, PRIMER\_NOMBRE, SEGUNDO\_NOMBRE) Dos nombres y Dos apellidos que corresponden generalmente a la identificación de la persona en el registro, que colocan los padres en el momento de registrar a su hijo(a).

**Fecha de nacimiento:** (FECHA\_NACIMIENTO) Es la fecha en la que aparece registrado el nacimiento de una persona. Esta variable relacionada con las variables de nombres y apellidos, constituye una excelente llave de identificación de las personas.

### 5.1.2. Variables de Ubicación

**Dirección:** (DIRECCION) Lugar reportado en el registro administrativo.

**Teléfono:** (TELEFONO) El número de teléfono tanto móvil como fijo de los residentes para su ubicación.

**Correo electrónico:** (CORREO\_ELECTRONICO) el correo electrónico permite enviar y recibir mensajes desde un navegador web.

**Lugar de residencia:** (ID\_MUN\_RESIDENCIA, COD\_MUN\_RESIDENCIA) corresponde al código del departamento como identificador y el código del municipio<sup>1</sup> de la División Político Administrativa de Colombia - DIVIPOLA<sup>2</sup>.

### 5.1.3. Variables Demográficas

**Lugar de nacimiento:** (ID\_MUN\_NACIMIENTO).

**País de nacimiento:** (ID\_PAIS\_NACIMIENTO)

**Sexo:** (ID\_SEXO).

**Estado civil:** (ID\_ESTADO\_CIVIL)

**Parentesco:** (ID\_PARENTESCO)

**Fecha de Defunción:** (FECHA\_DEFUNCION)

**Edad:** (EDAD)

### 5.1.4. Derivación de nuevas variables

**Estado de la Persona en el Registro:** (ID\_ESTADO\_PERSONA) Identifica el estado de la persona que se encuentra en el registro. Los estados se encuentran definidos en el Anexo 2 Tablas de Referencia.

---

<sup>1</sup> La información proveniente de las entidades promotoras de salud (EPS) es considerada como la de mejor calidad, completitud y actualización para la variable de municipio de residencia.

<sup>2</sup> Ver [Geovisor de Consulta de Codificación de la Divipola](#)

- ACTIVO: (1) Se asigna a las personas que son residentes vivos, porque se encuentran registradas en alguno de los registros administrativos evaluados en el desarrollo del proyecto. Todos los registros nuevos, a excepción de los que tienen como proveedor de información el registro administrativo de defunciones inician como activos.
- INACTIVO POR EMIGRACION: (2) Una vez construido el registro estadístico base de población se evalúa el estado de la persona como emigrante hacia el exterior, para esto se toma en cuenta el registro de Migración Colombia, el cual determina el estado de esta persona. El cruce se realiza por tipo y número de documento; en una primera fase se estandarizan las categorías de tipo de documento que maneja la oficina de Migración Colombia con los tipos de documento que maneja el registro estadístico base de población. Seguido se procede al cruce por documento de la persona, y se actualiza el campo estado en el cual se contempla que una persona ha durado más de un año fuera del país. El registro de Migración Colombia no cuenta con el nombre de la persona imposibilitando la realización de cruces por campos alfabéticos.
- INACTIVO POR PROCESO ADMINISTRATIVO: (3) Esta marca se realiza a registros que se han detectados como no válidos, pero no se pueden eliminar.
- INACTIVO POR MUERTE: (4) Se marca con este estado las personas que se encuentran registradas en el registro de defunciones, esta marca también se tiene en el momento de integrar el primer registro de la integración para actualizar si es activo o inactivo por muerte y al finalizar la construcción del registro estadístico base de población, nuevamente se compara uno a uno para verificar si se encuentra fallecido.
- INACTIVO POR DUPLICADO DETERMINISTICO: (5) Después de realizar diferentes tipos de análisis de acuerdo a las características que se explican en el capítulo 3.2.3 del presente documento, se pueden encontrar personas registradas dos o más veces. En estos casos se marca un registro como ACTIVO y los demás registros como INACTIVO POR DUPLICADO DETERMINISTICO. Esta marca se realiza al finalizar la construcción del registro estadístico base de población.
- Los registros inactivos por duplicados, no deben tenerse en cuenta para la generación de estadísticas.
- INACTIVO POR CANCELACION EN RNEC: (7) Esta marca se realiza a los registros que tienen un documento en la base de la Registraduría Nacional del Estado Civil en estado Cancelado, es decir los que tienen estado mayor o igual a 21.
- INACTIVO POR INACTIVIDAD EN BDUA: (8) Esta marca se realiza a los registros que en la BDUA tienen un estado de afiliación diferente a Activo o Suspendido.

**Proveedor de Información:** (ID\_PROVEEDOR) Identificador del proveedor de información dentro del registro, se encuentran definidos en el Anexo 2 Tablas de Referencia.

**Fonético de Nombres y Apellidos:** (FON\_PRIMER\_APELLIDO, FON\_SEGUNDO\_APELLIDO, FON\_PRIMER\_NOMBRE, FON\_SEGUNDO\_NOMBRE): Se aplica la función a su fonema<sup>3</sup>, para identificar el nombre por el sonido que produce al ser pronunciado.

Para la identificación de duplicados por nombres, apellidos y fecha de nacimiento, se utilizó la función que convierte los nombres y apellidos a su fonema. Para identificar fonéticos, se desarrolló la función FN\_CONVIERTE\_FONETICO, la cual recibe una cadena de caracteres como parámetro, realiza la conversión de cada carácter y retorna su representación fonética.

Tabla 2. Ejemplo de conversión de carácter a fonético

ENTRADA	SALIDA
YOHANA	YN
JOANA	YN
JENNY	YNI
YENY	YNI
GONZALES	GNZLZ
GONZALEZ	GNZLZ

**Conteo de Letras:** (ID\_CONTEO\_LETRAS) se genera una función que cuenta las letras del alfabeto contenidas en una cadena, con el propósito de identificar diferencias en nombres. Recibe como argumento de entrada una cadena de caracteres y retorna una cadena de 56 dígitos con el resultado del conteo donde los primeros 2 dígitos corresponden el total de la letra 'A', los siguientes 2 dígitos corresponden a la letra 'B' y así sucesivamente

Ej.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	
BERENICEHEREDIA	0	1	0	1	0	1	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

**Consistencia de los Campos de Identificación:** (CONSIST\_TIP\_IDE, CONSIST\_NUM\_IDE, CONSIST\_PRI\_NOM, CONSIST\_SEG\_NOM, CONSIST\_PRI\_APE, CONSIST\_SEG\_APE, CONSIST\_FECHA\_NAC, CONSIST\_SEXO) Es una marca que se establece para las principales variables de identificación y se definen de acuerdo con el siguiente criterio:

- 1 Consistente
- 1 Faltante
- 0 Inconsistente

<sup>3</sup> Unidad fonológica mínima que resulta de la abstracción o descripción teórica de los sonidos de la lengua

	<b>Metodología general para la construcción y actualización del Registro Estadístico Base de Población - REBP</b>	CÓDIGO: VERSIÓN: 01
PROCESO: Misional	SUBPROCESO: Diseño	

**Marcación de duplicado por tipo y número de identificación:**

(DUP\_DOC). Si dos o más registros se encuentran duplicados por tipo y número de documento uno de los registros se marca como 1 Consistente y los demás se marcan como 0 Duplicado(s).

(DUP\_DOC\_INI): Si dos o más registros se encuentran duplicados por tipo y número de documento se marcan todos los registros como 0 Duplicado(s).

**Marcación de duplicado por fonético de nombres, apellidos y fecha nacimiento:**

(DUP\_NOM): Si dos o más registros se encuentran duplicados por el fonético de los nombres y apellidos, junto con la fecha de nacimiento, uno de los registros se marca como 1 Consistente y los demás se marcan como 0 Duplicado(s).

(DUP\_NOM\_INI): Si dos o más registros se encuentran duplicados por el fonético de los nombres y apellidos, junto con la fecha de nacimiento, se marcan todos los registros como 0 Duplicado(s).

**Control de registros administrativos con los que cruza la persona:**

(CRUZO\_DOC y CRUZO\_NOM): Si se encuentra la persona en un Registro Administrativo se graba el código del proveedor del registro administrativo hasta conformar una cadena con los códigos de todos los registros con los que cruza, por ejemplo, si cruza por SISBEN, RUV Y BDUA QUEDA: 020809

**5.2 Codificación, homologación y validación para cada registro administrativo fuente**

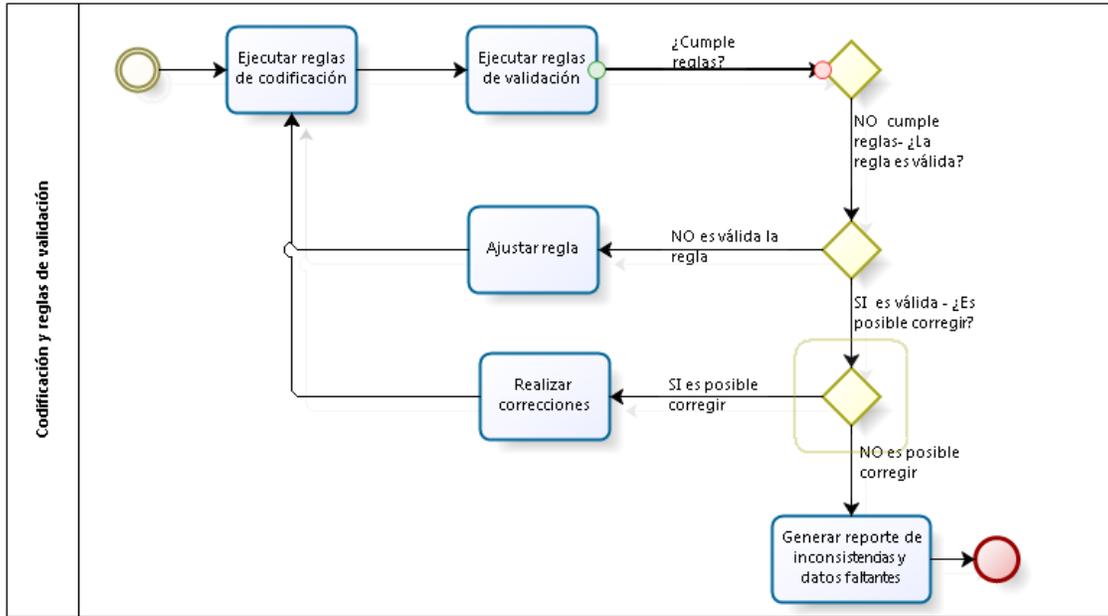
Se aplican un conjunto de normas para revisar la coherencia de la información del registro administrativo recibido y se establecen las medidas de corrección más apropiadas, acordes con los objetivos y metas del REBP. Los criterios de validación se caracterizan por:

- Respetar al máximo la información correctamente registrada en cada variable.
- Aplicar universalmente los criterios a todos los registros.
- Emplear reglas de congruencia lógica.
- Asignar valores sólo si existe información sustentada en otras variables asociadas.
- No eliminar casos que pudieran ser considerados erróneos.

Debido a que los distintos proveedores de información pueden enviar periódicamente archivos con nombres de variables diferentes y con diferente estructura, se determine la referencia entre los nombres de variables que se utilizan en el registro estadístico base de población REBP y los nombres de variables de los archivos de cada proveedor de información.

En esta etapa del proceso se crean reglas de validación para detectar inconsistencias en los datos, esto permite identificar problemas en la información tales como discrepancias respecto a un valor de referencia, valores atípicos, falta de respuesta en las variables y errores de codificación.

Figura 3. Codificación y Reglas de validación



En la figura 3 se describen las validaciones necesarias que permitan garantizar que la información entregada por los proveedores contenga la calidad y el nivel de consistencia requerido. El conjunto de reglas deberá ser revisado con periodicidad anual para determinar si la regla permanece vigente o debe modificarse.

Existen dos tipos de validaciones que se realizan:

- Validaciones preliminares que consisten en eliminar caracteres especiales revisando codificación y longitud de las variables, separación de nombres y apellidos.
- Validación de consistencias aplicadas a las variables de identificación<sup>4</sup>. Las inconsistencias serán reportadas para cada registro administrativo por medio de una ficha<sup>5</sup> estadística donde se registra la presencia de datos y la consistencia de registros por variables (Anexo 3).

<sup>4</sup> Ver 5.1.4 *Consistencia de los Campos de Identificación*

<sup>5</sup> Contiene el diccionario de las variables, los datos básicos de la información recibida y el resultado de las validaciones.

	<b>Metodología general para la construcción y actualización del Registro Estadístico Base de Población - REBP</b>	CÓDIGO: VERSIÓN: 01
PROCESO: Misional	SUBPROCESO: Diseño	

### 5.2.1 Depuración e imputación de la variable sexo

En este subproceso se completan valores faltantes indispensables en el análisis demográfico, imputando la variable sexo de acuerdo con la mayor frecuencia que presentan el primer nombre por sexo

### 5.2.2 Definición de duplicados por análisis determinístico

Se identifican y se marcan en el registro estadístico base de población, así como en los registros administrativos, la información repetida teniendo en cuenta los siguientes criterios:

- Identificación de los registros duplicados por tipo y número de identificación<sup>6</sup>.
- Identificación de los registros duplicados por el fonético del primer apellido, segundo apellido, primer nombre, segundo nombre y fecha de nacimiento<sup>7</sup>.

### 5.3 Emparejamiento, integración y mejoramiento de variables del REBP

Como punto de partida para la consolidación del REBP, se inicia con el registro administrativo con la mayor cantidad de observaciones que corresponde a la RNEC, cuya entidad responsable, es la encargada de emitir los documentos de identificación nacionales. De esta forma, se establece el orden de integración, iniciando por el registro administrativo catalogado con mejor calidad. Posteriormente, se incluyen las restantes fuentes de información a través del emparejamiento<sup>8</sup> de registros mediante las variables vínculo, que corresponden a la identificación de las personas.

Los identificadores pueden no ser únicos por ejemplo en el caso de Colombia debido a la vigencia de diferentes tipos y números de documento para diferentes personas, a los errores en el registro de los identificadores en las diferentes bases de datos o la homonimia de los nombres.

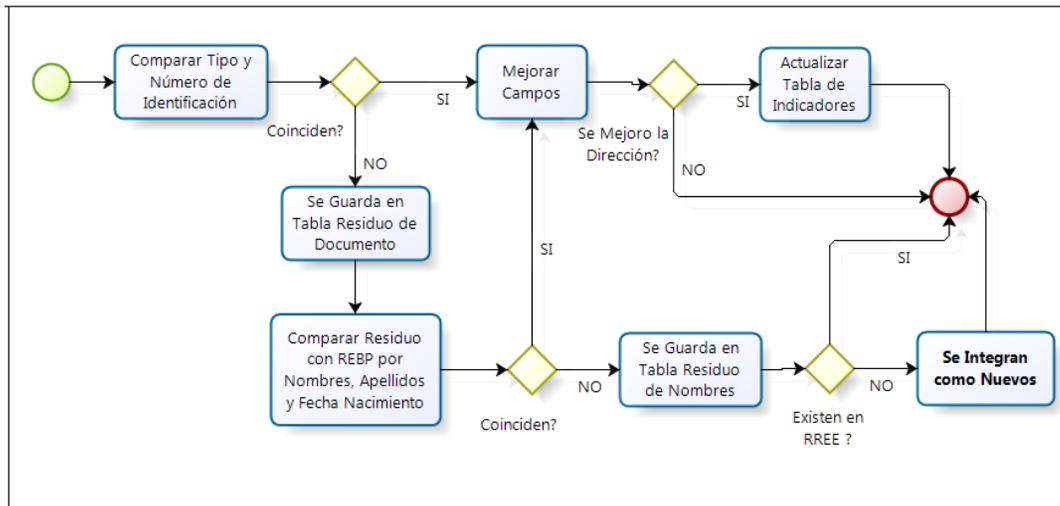
En la Figura 4, se define el proceso de emparejamiento entre el registro administrativo del proveedor de información y el registro estadístico base de población para el mejoramiento de las variables. El proceso de mejoramiento busca aprovechar la información consistente en un registro administrativo o base a integrar para mejorar campos inconsistentes, vacíos o que no son coherentes con las variables que representan en el registro estadístico base de población en construcción.

<sup>6</sup> Ver 5.1.4 Marcación de duplicado por tipo y número de identificación

<sup>7</sup> Ver 5.1.4 Marcación de duplicado por fonético de nombres, apellidos y fecha nacimiento

<sup>8</sup> El emparejamiento de registros se refiere a la tarea de comparar registros que corresponden a la misma persona en dos o más archivos e integrarlos cuando estos coinciden.

Figura 4. Proceso de Emparejamiento e Integración de Registros



El mejoramiento se realiza en dos fases secuenciales tal como se muestra en la Figura 4:

1. **Registros que emparejan por tipo y número de documento:** Se utilizan para actualizar nombres, apellidos, fechas de nacimiento, sexo, teléfono, dirección, correo electrónico y municipio de residencia. Este emparejamiento se realiza únicamente con registros que tienen tipo y número de documento marcado como consistente.
2. **Registros que no se emparejan por tipo y número de documento,** es decir el residuo del paso anterior se emparejan por nombres, apellidos y fecha de nacimiento y se utilizan para actualizar tipo y número de documento, sexo, teléfono, dirección, correo electrónico y municipio de residencia. Para este proceso de emparejamiento se utilizó la función de comparación de letras<sup>9</sup>

Los registros del proveedor de información que no coincidieron en la comparación por tipo y número de identificación, ni tampoco coincidieron por nombres, apellidos y fecha de nacimiento se integran como nuevos al registro estadístico de población y a la tabla de identificadores.

El proceso de integración y mejoramiento de registros cuenta con dos procedimientos denominados PR\_INTEGRACION\_PRIMER\_RA y PR\_INTEGRACION, que son los más importantes; y el eje sobre el

<sup>9</sup> Ver 5.1.4 *Conteo de Letras*

cual se realizan todas las mejoras en el registro estadístico base de población, así como la integración de los nuevos registros. Este procedimiento automatiza el proceso de integración de registros. Las tareas que realiza PR\_INTEGRACION son integrar registros, mejoramiento de registros y trazabilidad del id tabla con la dirección que es almacenada en una tabla de identificadores. Adicionalmente, esta función llama otras funciones desarrolladas para la validación y estandarización de la información. Al finalizar el proceso de integración se realizan las siguientes marcaciones en el campo ID\_ESTADO\_PERSONA<sup>10</sup>.

---

<sup>10</sup> Ver 5.4.1 *Estado de la Persona en el Registro*

	<b>Metodología general para la construcción y actualización del Registro Estadístico Base de Población - REBP</b>	CÓDIGO: VERSIÓN: 01
PROCESO: Misional	SUBPROCESO: Diseño	

## 6. ANÁLISIS Y VALIDACIÓN DE RESULTADOS

En esta etapa del proceso se aprovecha que la temporalidad del REBP coincide con el Censo Nacional de Población y Vivienda 2018 (CNPV), lo que permite el contraste de estructuras poblacionales, indicadores demográficos básicos y volúmenes poblacionales a niveles subnacionales. Adicionalmente, gracias a que el CNPV cuenta con la identificación de las personas, mediante el emparejamiento determinístico por tipo y número de identificación, es posible comparar la cobertura del REBP respecto a la información censal.

Se realizan dos procesos. El primero consiste en la verificación y consistencia de la base de datos integrada, tomando las variables de sexo y edad de la persona con estado activo dentro del REBP, contrastando las estructuras demográficas a nivel nacional, departamental y municipal con los resultados provenientes del CNPV.

La segunda consiste en la generación de diferentes indicadores para evaluar la calidad de reporte de las variables que hacen parte del REBP. En este proceso, se obtienen los principales indicadores demográficos de la base integrada, comparándolos con la fuente censal. Dentro de los indicadores demográficos se encuentran:

- Índice de masculinidad
- Razón de dependencia demográfica.
- Índice de envejecimiento
- Relación de niños por mujer

En esta fase se verifica la coherencia y la consistencia de la información, que permitan garantizar la calidad de los productos estadísticos.

	<b>Metodología general para la construcción y actualización del Registro Estadístico Base de Población - REBP</b>	CÓDIGO: VERSIÓN: 01
PROCESO: Misional	SUBPROCESO: Diseño	

## 7. PRODUCTOS DE DIFUSIÓN

Las anteriores etapas describen la obtención y generación de información demográfica proveniente de fuentes no convencionales, cuya característica principal radica en la posibilidad de realizar contrastes con los censos, siendo estos, la fuente tradicional para dichos fines. Partiendo de la base integrada del REBP como insumo para producción de estudios demográficos, al igual que los provenientes de la base censal, se amplía el conocimiento de las dinámicas poblacionales en Colombia. Se busca que el REBP tenga un exhaustivo aprovechamiento por parte de las entidades encargadas del diseño y seguimiento de la política pública, la comunidad académica y la sociedad en general, configurándose en un bien público regional con la capacidad de ofrecer información poblacional en periodos intercensales.

Con la información generada se cuentan con insumos suficientes para la elaboración de:

- Informes técnicos.
- Herramientas interactivas de visualización.
- Webinar de presentación de resultados.
- Estudios demográficos basados en registros administrativos.
- Divulgación y socialización a los institutos nacionales de estadística de la región.
- Comunicados de prensa.

	<b>Metodología general para la construcción y actualización del Registro Estadístico Base de Población - REBP</b>	CÓDIGO: VERSIÓN: 01
PROCESO: Misional	SUBPROCESO: Diseño	

## 8. GLOSARIO

**Base de datos:** Son usadas para almacenar microdatos estadísticos consisten en una serie de tablas de bases de datos relacionadas.

**Consistencia:** Se considera que dos encuestas son consistentes si están relacionadas con la misma población y si los objetos tienen los mismos valores para variables comunes, con las mismas estimaciones.

**Datos agregados:** Datos relacionados con sumas de datos, valores resumen de los datos, macrodatos, que hacen referencia a grupos de objetos y normalmente se presentan como contenido en tablas estadísticas o tabulados.

**Duplicidad:** Se considera duplicidad cuando son detectados dos registros con la misma información en las variables básicas.

**Emparejamiento:** Comparación de los vínculos entre dos o más registros. El resultado es una coincidencia o una no coincidencia (no emparejamiento). También se le conoce como emparejamiento exacto.

**Estadísticas oficiales:** Es la información de uso obligatorio por parte de las entidades del Estado en documentos de política, planes, programas, proyectos y para la transmisión de información del país a organismos internacionales. Debe cumplir los siguientes requisitos:

1. Que la operación estadística (de acuerdo con el Decreto 1743 de 2016 en su artículo 1, una operación estadística es "la aplicación de un proceso estadístico sobre un objeto de estudio que conduce a la producción de información estadística"), que la genera esté incorporada en el Plan Estadístico Nacional.
2. Que la operación estadística que la genera haya aprobado la evaluación de la calidad estadística establecido para el SEN.

**Estandarización:** La Estandarización permite la creación de normas o estándares que establecen las características comunes con las que deben cumplir los productos

**Ficha estadística del registro administrativo:** Formato en el cual se reportan las inconsistencias encontradas para cada registro administrativo, así como también el registro de la presencia de datos y la consistencia de registros por variables.

**Imputación:** Los valores perdidos se reemplazan por valores imputados, que se calculan usando una distribución de probabilidades o un modelo determinístico.

	<b>Metodología general para la construcción y actualización del Registro Estadístico Base de Población - REBP</b>	CÓDIGO: VERSIÓN: 01
PROCESO: Misional	SUBPROCESO: Diseño	

**Indicadores sociodemográficos:** Son relaciones estadísticas referidas a algún tema en particular, por ejemplo, la natalidad, la mortalidad o la fecundidad de una población específica y en un momento determinado

**Integración:** Es el proceso en el cual se combinan datos de muchas fuentes. Durante esta fase se crean la población del registro y los objetos derivados, se importan las variables de diferentes fuentes y se generan variables derivadas.

**Inventario de registros:** Comprende el listado consolidado de registros administrativos, suministrado por entidades externas y en el cual se incorpora información relacionada a la temporalidad de la información y principales variables.

**Metadato:** Información necesaria para usar e interpretar las estadísticas. Describe los datos mediante definiciones de las poblaciones, objetos, variables, metodología y calidad.

**Operación estadística:** Conjunto de procesos y actividades que, partiendo de la recolección sistemática de datos, conduce a la producción de resultados agregados.

**Plan Estadístico Nacional:** Es la hoja de ruta que tendrá el país para la producción estadística para el periodo comprendido entre 2017 – 2022. Este documento guiará a las entidades del Sistema Estadístico Nacional en el fortalecimiento de sus actividades estadísticas, indicando estrategias y acciones que permitan, entre otros propósitos, la innovación en la producción y difusión de estadísticas oficiales, el fortalecimiento y mayor aprovechamiento de los registros administrativos como fuentes alternativas para la generación de estadísticas oficiales, así como elevar la calidad estadística.

**Referentes internacionales:** Mecanismo por el cual diferentes organismos internacionales han trabajado en el desarrollo de lineamientos que permitan mejorar la calidad en la producción estadística.

**Registro:** Corresponde a una lista de objetos que busca contener el total de elementos de una población específica y datos individualizados sobre sus principales características. Adicionalmente, permiten identificar de manera inequívoca a cada objeto de tal forma que el registro pueda ser actualizado y ampliado con nuevas variables para cada objeto.

**Registro Administrativo (RRAA):** Es el registro utilizado con fines administrativos en un sistema de información administrativo. Contendrá los objetos por administrar, incluyendo la identificación y variables usadas para propósitos administrativos (Wallgreen & Wallgreen, 2012).

**Registro Estadístico Base de Población (REBP):** Se crea a partir de la transformación de uno o varios registros administrativos, de tal manera que los objetos y las variables se ajusten para satisfacer las necesidades estadísticas (Wallgreen & Wallgreen, 2012). También es posible crear un registro estadístico a partir del procesamiento de registros administrativos junto con otros registros estadísticos (Decisión 780 CAN. 2013. Pg. 2).

**Sistemas de información geográfica:** Es un conjunto de herramientas que integra y relaciona diversos componentes (usuarios, hardware, software, procesos) que permiten la organización, almacenamiento, manipulación, análisis y modelización de grandes cantidades de datos procedentes del mundo real que están vinculados a una referencia espacial, facilitando la incorporación de aspectos sociales-culturales, económicos y ambientales que conducen a la toma de decisiones de una manera más eficaz.

**Sistema Estadístico Nacional:** Conjunto articulado de componentes, que, de manera organizada y sistemática, garantiza la producción y difusión de las estadísticas oficiales a nivel nacional y territorial que requiere el país. Sus componentes son las entidades u organizaciones que lo integran, usuarios, procesos e instrumentos técnicos para la coordinación, políticas, principios, fuentes de información, infraestructura tecnológica y talento humano.

**Validación:** Proceso de revisión con el propósito de verificar si cumple con las especificaciones esperadas.

**Variable:** Atributo medible de un objeto.

	<b>Metodología general para la construcción y actualización del Registro Estadístico Base de Población - REBP</b>	CÓDIGO: VERSIÓN: 01
PROCESO: Misional	SUBPROCESO: Diseño	

## 9. BIBLIOGRAFÍA

Departamento Nacional de Estadística. Decreto 1743 de 2016. Recuperado de: [https://www.dane.gov.co/files/sen/normatividad/decreto\\_1743\\_noviembre\\_1\\_2016.pdf](https://www.dane.gov.co/files/sen/normatividad/decreto_1743_noviembre_1_2016.pdf)

Departamento Nacional de Estadística. 2009. Plan de Fortalecimiento de Registros Administrativos. Estrategias para el fortalecimiento estadístico territorial.

Departamento Nacional de Estadística. Plan Estadístico Nacional 2017-2022

Departamento Nacional de Estadística- Planes Indicativos. Seguimiento Plan de Acción. Recuperado de: [https://www.dane.gov.co/files/control\\_participacion/planes\\_institucionales/Planes\\_Indicativos/Seguimiento\\_plandeaccion\\_IItri\\_2019.pdf](https://www.dane.gov.co/files/control_participacion/planes_institucionales/Planes_Indicativos/Seguimiento_plandeaccion_IItri_2019.pdf)

Departamento Nacional de Estadística. Sistema Estadístico Nacional. Recuperado de: <https://www.dane.gov.co/index.php/sistema-estadistico-nacional-sen>

Decisión 780. 2013. Creación y Actualización del Sistema de Registros Estadísticos en los Países Miembros de la Comunidad Andina (CAN).

Decreto 1742 de 2016. Artículo 2.2.3.2.1. Condiciones y características de la estadística oficial.

Documento donde se listan y detallan las variables mínimas a incluir en el registro de personas, para cada unidad estadística. Recuperado de: [http://estadisticas.comunidadandina.org/eportal/contenidos/imagenes/file/reunioncan/registros\\_7\\_Propuesta%20sobre%20Registro%20de%20Personas%20-%20SGCAN.doc](http://estadisticas.comunidadandina.org/eportal/contenidos/imagenes/file/reunioncan/registros_7_Propuesta%20sobre%20Registro%20de%20Personas%20-%20SGCAN.doc)

Estandarización (s.f.). En *Gobierno de Mexico*. Recuperado el 20 de agosto de 2019 de <https://www.gob.mx/se/articulos/que-es-la-estandarizacion>

Fellegi, I. &. (1969). A theory for Record Linkage. *Journal of the American Statistical Association*, 64(328), 1183-1210.

Heredía, B., Rodríguez, S., & Mendoza, A. (2015). *METODOLOGÍA PARA LA CONSTRUCCIÓN DEL REGISTRO ESTADÍSTICO DE SAN ANDRES, PROVIDENCIA Y SANTA CATALINA*. Bogotá: DANE

Indicadores demográficos. (s.f.). En *Contenidos digitales*. Recuperado el 20 de agosto de 2019 de [http://contenidosdigitales.ulp.edu.ar/exe/geo-olitica/los\\_indicadores\\_demogrficos.html](http://contenidosdigitales.ulp.edu.ar/exe/geo-olitica/los_indicadores_demogrficos.html)

R Core Team. (2015). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

Wallgren A. & Wallgren B. 2012. Estadísticas basadas en registros, Aprovechamiento estadístico de datos administrativos. INEGI.

## 10. ANEXOS

### Anexo 1. Diccionario de datos

ORDEN VARIABLE EN BASE	NOMBRE VARIABLE	DESCRIPCIÓN VARIABLE	OBSERVACIONES	CATEGORÍAS	TIPO DE DATOS	TAMANO DATO
1	ID_TABLA	Consecutivo en la tabla			numérico	9
2	ID_ESTADISTICO	Identificador en el registro administrativo	Este identificador se toma del registro administrativo que lo crea		numérico	9
3	ID_TIPO_IDENTIFICACION	Tipo de Identificación		1 Registro Civil 2 Tasa de Identidad	numérico	2
4	NUMERO_IDENTIFICACION	Número de Identificación			varchar	20
5	PRIMER_APELLIDO	Primer Apellido			varchar	60
6	SEGUNDO_APELLIDO	Segundo Apellido			varchar	60
7	PRIMER_NOMBRE	Primer Nombre			varchar	60
8	SEGUNDO_NOMBRE	Segundo Nombre			varchar	60
9	ID_SEXO	Sexo		1 Hombre 2 Mujer	numérico	2
10	FECHA_NACIMIENTO	Fecha de Nacimiento			datetime	
11	EDAD	Edad de la persona	Calculada al momento de la integración del sistema		numérico	3
12	DIRECCION	Lugar de residencia habitual de la persona en el país			varchar	300
13	TELEFONO	Teléfono de contacto			varchar	50
14	CORREO_ELECTRONICO	Correo Electronico			varchar	300
15	COD_MUN_RESIDENCIA	Código Municipio de Residencia	Código de la DIVIPOLA		varchar	5
16	ID_MUN_RESIDENCIA	Municipio de Residencia	Identificador asociado al código de la DIVIPOLA		numérico	4
17	ID_MUN_NACIMIENTO	Municipio de Nacimiento	Identificador asociado al código de la DIVIPOLA		numérico	4
18	ID_PAIS_NACIMIENTO	País de Nacimiento	Identificador asociado a la tabla de países		numérico	4
19	ID_CLASE	Clase	El dato original tiene estas categorías: 1 Cabecera Municipal	1 Cabecera Municipal 2 Centro Poblado	numérico	2
20	ID_CENTRO_POBLADO	Centro Poblado			numérico	2
21	ID_ESTADO_CIVIL	Estado civil	El dato original tiene estas categorías: 1 Casado(a)	1 Soltero 2 Casado	numérico	2
22	ID_PARENTESCO	Parentesco con el cabeza de familia o jefe	El dato original tiene estas categorías: 1 Jefe o cabeza del hogar		numérico	2
23	FECHA_DEFUNCION	Fecha del fallecimiento			datetime	
24	ID_ESTADO_PERSONA	Estado actual de la persona en el registro		1 Activo	numérico	1
25	ID_PROVEEDOR	Proveedor de información		1 REGISTRO SAN ANDRÉS 2 CENSA	numérico	2
26	FON_PRIMER_APELLIDO	Fonético Primer Apellido			varchar	50
27	FON_SEGUNDO_APELLIDO	Fonético Segundo Apellido			varchar	50
28	FON_PRIMER_NOMBRE	Fonético Primer Nombre			varchar	50
29	FON_SEGUNDO_NOMBRE	Fonético Segundo Nombre			varchar	50
30	CONSIST_TIP_IDE	Consistencia del Tipo de Identificación		1 Consistente	numérico	1
31	CONSIST_NUM_IDE	Consistencia del número de identificación		0 NULL Inconsistente	numérico	1
32	CONSIST_PRI_APE	Consistencia del Primer Apellido		0 NULL Inconsistente	numérico	1
33	CONSIST_SEG_APE	Consistencia del Segundo Apellido		0 NULL Inconsistente 1 Consistente	numérico	1
34	CONSIST_PRI_NOM	Consistencia del Primer Nombre		0 NULL Inconsistente	numérico	1
35	CONSIST_SEG_NOM	Consistencia del Segundo Nombre		0 NULL Inconsistente 1 Consistente	numérico	1
36	CONSIST_FEC_NAC	Consistencia de la Fecha de Nacimiento		0 NULL Inconsistente 1 Consistente	numérico	1
37	CONSIST_SEXO	Consistencia del Sexo		0 NULL Inconsistente 1 Consistente	numérico	1
38	DUP_DOC_INI	Duplicado por Documento	Se cuentan todos los registros duplicados	1 Duplicado	numérico	1
39	DUP_NOM_INI	Duplicado por Nombres y Fecha de nacimiento	Se cuentan todos los registros duplicados	1 Duplicado	numérico	1
40	DUP_DOC	Duplicado por Documento	Se cuentan todos los registros duplicados, menos uno	0 Duplicado	numérico	1
41	DUP_NOM	Duplicado por Nombres y Fecha de nacimiento	Se cuentan todos los registros duplicados, menos uno	0 Duplicado	numérico	1
42	ID_CONTEO_LETRAS	Vector que tiene la cantidad de letras que cada letra de la que se especifica con el código de la variable de cruce	De esta variable en adelante se encuentran las variables asociadas al CENSA 2019	0 Duplicado	varchar	100
43	CRUZO_DOC	CADA UNA DE LAS QUE SE ESPECIFICA CON EL CÓDIGO DE LA VARIABLE DE CRUCE			varchar	100
44	CRUZO_NOM	CADA UNA DE LAS QUE SE ESPECIFICA CON EL CÓDIGO DE LA VARIABLE DE CRUCE			varchar	100
45	UID_ORIGEN	Modo de Recolección				
46	COD_ENCUESTAS	Identificador de la Encuesta				
47	ID_VIVIENDA	Identificador de la Vivienda				
48	ID_HOGAR	Identificador del hogar				
49	ID_PERSONA_HOGAR	Identificador de la persona en el hogar con código DIVIPOLA del municipio de residencia				
50	U_MUNICIPIO	Municipio de Residencia				
51	UVA_DIRUND	Dirección Residencia				
52	UA_CLASE	Clase				
53	UA2_CPOB	Centro Poblado				
54	UA1_LOCALIDAD	Localidad				
55	U_CO	Centro Operativa				
56	U_AO	Area Operativa				
57	U_UC	Unidad de Cobertura				

**Validaciones de variables principales**

VARIABLE	VALIDACIÓN		
Tipo de Identificación		RC	8 ó 11 dígitos – Numérico 10 – Alfanumérico o Numérico
		TI	10 – Numérico 11 dígitos - Numérico
		CC	3-8 ó 10 dígitos - Numérico
		CE	3-6 dígitos – Numérico
		CD	9 ó 11 – Alfanumérico
		PA	< 17 dígitos – Alfanumérico
		NI	3-11 - Numérico
Número de Identificación	Validación por tipo de identificación y por longitud		
Fecha de nacimiento	Valor tipo fecha	Fecha de nacimiento menor a la fecha de registro Al calcular edad < 110 años Diferente a Vacío	
Nombres y Apellidos	No numéricos Mayor a 2 caracteres No debe contener cadenas como NN, NA No debe contener caracteres especiales		
Edad	< 110 años		
Sexo	Tabla de referencia SEXO		
Municipio de Nacimiento	Tabla de referencia MUNICIPIOS – DIVIPOLA		
Municipio de Residencia			
Departamento de Nacimiento	Tabla de referencia DEPARTAMENTOS – DIVIPOLA		
Departamento de Residencia			
País de Nacimiento	Tabla de referencia PAISES		
Clase	Tabla de referencia CLASES		
Centro Poblado	Tabla de referencia CENTROS_POBLADOS		

Estado civil	Tabla de referencia ESTADOS_CIVILES	
Parentesco	Tabla de referencia PARENTESCOS	
Estado Persona	Tabla de referencia ESTADOS_PERSONA	

**Anexo 2. Tablas de Referencia**

Se presentan las tablas de referencia que se diseñaron en la base de datos:

**Tipos de Identificación - TIPOS\_IDENTIFICACION**

NOMBRE	DESCRIPCION	TIPO	LONG. MAX.
ID_TIPO_IDENTIFICACION	Identificador del Tipo de Identificación	SMALLINT	
COD_TIPO_IDENTIFICACION	Código del Tipo de Identificación	VARCHAR	2
NOMBRE_TIPO_IDENTIFICACION	Nombre del Tipo de Identificación	VARCHAR	50

ID_TIPO_IDENTIFICACION	COD_TIPO_IDENTIFICACION	NOMBRE_TIPO_IDENTIFICACION
1	RC	REGISTRO CIVIL
2	TI	TARJETA DE IDENTIDAD
3	CC	CEDULA DE CIUDADANIA
4	CE	CEDULA DE EXTRANJERIA
5	PA	PASAPORTE
6	CD	CARNET DIPLOMATICO
7	NI	NUMERO DE IDENTIFICACION TRIBUTARIA
8	MS	MENOR SIN DOCUMENTO DE IDENTIFICACION
9	AS	ADULTO SIN DOCUMENTO DE IDENTIFICACION

**Sexo – SEXO**

NOMBRE	DESCRIPCION	TIPO	LONG. MAX.
ID_SEXO	Identificador del Sexo	SMALLINT	
COD_SEXO	Código del Sexo	VARCHAR	1

NOMBRE	DESCRIPCION	TIPO	LONG. MAX.
NOMBRE_SEXO	Nombre del Sexo	VARCHAR	50

ID_SEXO	COD_SEXO	NOMBRE_SEXO
1	M	HOMBRE
2	F	MUJER

**Clases – CLASES**

NOMBRE	DESCRIPCION	TIPO	LONG. MAX.
ID_CLASE	Identificador de la Clase	SMALLINT	
NOMBRE_CLASE	Nombre de la Clase	VARCHAR	20

ID_CLASE	NOMBRE_CLASE
1	CABECERA MUNICIPAL
2	CENTRO POBLADO
3	RURAL DISPERSO

**Estados Civiles - ESTADOS\_CIVILES**

NOMBRE	DESCRIPCION	TIPO	LONG. MAX.
ID_ESTADO_CIVIL	Identificador del Estado Civil	SMALLINT	
NOMBRE_ESTADO_CIVIL	Nombre del Estado Civil	VARCHAR	20

ID_ESTADO_CIVIL	NOMBRE_ESTADO_CIVIL
1	CASADO
2	UNION LIBRE
3	DIVORCIADO
4	SEPARADO
5	VIUDO
6	SOLTERO

**Parentescos – PARENTESCOS**

NOMBRE	DESCRIPCION	TIPO	LONG. MAX.
ID_PARENTESCO	Identificador del Parentesco	SMALLINT	
NOMBRE_PARENTESCO	Nombre del Parentesco	VARCHAR	50

ID_PARENTESCO	NOMBRE_PARENTESCO
1	JEFE O CABEZA DEL HOGAR
2	PAREJA (CONYUGE, COMPAÑERO, ESPOSO)
3	HIJO
4	HIJASTRO
5	YERNO, NUERA
6	PADRE, MADRE
7	SUEGRO
8	HERMANO, HERMANASTRO
9	CUÑADO
10	NIETO
11	TIO
12	ABUELO
13	OTRO PARIENTE
14	EMPLEADO DOMESTICO
15	OTRO NO PARIENTE

**Estado de la Persona en el Registro - ESTADOS\_PERSONA**

NOMBRE	DESCRIPCION	TIPO	LONG. MAX.
ID_ESTADO_PERSONA	Identificador del Estado de la Persona	SMALLINT	
NOMBRE_ESTADO_PERSONA	Nombre del Estado de la Persona	VARCHAR	50

ID_ESTADO_PERSONA	NOMBRE_ESTADO_PERSONA
1	ACTIVO
2	INACTIVO POR EMIGRACION
3	ELIMINADO POR PROCESO ADMINISTRATIVO
4	INACTIVO POR MUERTE
5	INACTIVO POR DUPLICADO DETERMINISTICO
6	INACTIVO POR DUPLICADO PROBABILISTICO
7	INACTIVO POR CANCELACION EN RNEC
8	INACTIVO POR INACTIVIDAD EN BDU A

**Proveedores de la Información - PROVEEDORES\_INFORMACION**

NOMBRE	DESCRIPCION	TIPO	LONG. MAX.
ID_PROVEEDOR_INFORMACION	Identificador del Proveedor de la Información	SMALLINT	
NOMBRE_PROVEEDOR_INFORMACION	Nombre del Proveedor de la Información	VARCHAR	100

ID_PROVEEDOR_INFORMACION	NOMBRE_PROVEEDOR_INFORMACION
2	SISBEN
3	NACIMIENTOS
4	DEFUNCIONES
5	EDUCACION
6	LEA
7	PILA
8	RUV
9	BDUA
10	RUT
11	NUEVA EPS
12	SANITAS EPS
13	SIGEP
14	SNIES
15	SIET
....	.....

**Departamentos (DIVIPOLA) – DEPARTAMENTOS**

NOMBRE	DESCRIPCION	TIPO	LONG. MAX.
ID_DEPARTAMENTO	Identificador del Departamento	SMALLINT	
COD_DEPARTAMENTO	Código del Departamento	VARCHAR	2
NOMBRE_DEPARTAMENTO	Nombre del Departamento	VARCHAR	100

ID_DEPARTAMENTO	COD_DEPARTAMENTO	NOMBRE_DEPARTAMENTO
1	05	ANTIOQUIA
2	08	ATLANTICO
3	11	BOGOTA, D.C.
4	13	BOLIVAR
5	15	BOYACA

ID_DEPARTAMENTO	COD_DEPARTAMENTO	NOMBRE_DEPARTAMENTO
6	17	CALDAS
7	18	CAQUETA
8	19	CAUCA
9	20	CESAR
10	23	CORDOBA
...	...	...

**Municipios (DIVIPOLA) – MUNICIPIOS**

NOMBRE	DESCRIPCION	TIPO	LONG. MAX.
ID_MUNICIPIO	Identificador del Municipio	SMALLINT	
COD_MUNICIPIO	Código del Municipio	VARCHAR	5
NOMBRE_MUNICIPIO	Nombre del Municipio	VARCHAR	100

ID_MUNICIPIO	COD_MUNICIPIO	NOMBRE_MUNICIPIO
1	05001	MEDELLIN
2	05002	ABEJORRAL
3	05004	ABRIAQUI
4	05021	ALEJANDRIA
5	05030	AMAGA
6	05031	AMALFI
7	05034	ANDES
8	05036	ANGELOPOLIS
9	05038	ANGOSTURA
10	05040	ANORI
...	...	...

**Centros Poblados - CENTROS\_POBLADOS**

NOMBRE	DESCRIPCION	TIPO	LONG. MAX.
ID_CENTRO_POBLADO	Identificador del Centro Poblado	SMALLINT	
COD_CENTRO_POBLADO	Código del Centro Poblado	VARCHAR	8
NOMBRE_CENTRO_POBLADO	Nombre del Centro Poblado	VARCHAR	100
CLASE	Clase del Centro Poblado	VARCHAR	10

ID_CENTRO_POBLADO	COD_CENTRO_POBLADO	NOMBRE_CENTRO_POBLADO	CLASE
1	05001000	MEDELLIN	CM
2	05001001	PALMITAS	C
3	05001004	SANTA ELENA	C
5	05001009	ALTAVISTA	C
6	05001010	AGUAS FRIAS	CP
7	05001012	LA LOMA	CP
8	05001013	SAN JOSE DEL MANZANILLO	CP
9	05001032	TRAVESIAS	CP
10	05002000	ABEJORRAL	CM
...	...	...	...

**Países - PAISES**

NOMBRE	DESCRIPCION	TIPO	LONG. MAX.
ID_PAIS	Identificador del País	SMALLINT	
COD_PAIS	Código del País	VARCHAR	3
NOMBRE_PAIS	Nombre del País	VARCHAR	100

ID_PAIS	COD_PAIS	NOMBRE_PAIS
1	004	AFGHANISTAN
2	008	ALBANIA
3	010	ANTARCTICA
4	012	ALGERIA
5	016	AMERICAN SAMOA
6	020	ANDORRA
7	024	ANGOLA
8	028	ANTIGUA AND BARBUDA
9	031	AZERBAIJAN
10	032	ARGENTINA
...	...	...

### Anexo 3. Ficha estadística

	<b>FICHA ESTADISTICA DE REGISTROS ADMINISTRATIVOS PROVENIENTES DE FUENTES EXTERNAS</b>	<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 30%;">CODIGO</td> <td></td> </tr> <tr> <td>VERSIÓN</td> <td></td> </tr> <tr> <td>PÁGINA</td> <td></td> </tr> <tr> <td>FECHA</td> <td></td> </tr> </table>	CODIGO		VERSIÓN		PÁGINA		FECHA	
CODIGO										
VERSIÓN										
PÁGINA										
FECHA										
<b>CARACTERISTICAS DE LA INFORMACIÓN</b>										
<b>ENTIDAD QUE SUMINISTRA LA INFORMACION</b>		DANE								
<b>NOMBRE DEL ARCHIVO ORIGINAL</b>		REG_EST_NCIONAL_2018.TXT								
<b>UBICACIÓN ARCHIVO ORIGINAL</b>		<a href="#">\\systema81\Regadmin\Registros Administrativos\Nacional 2018\Integrados</a>								
<b>TIPO DE ARCHIVO ORIGINAL</b>		TXT								
<b>NOMBRE DE ARCHIVO DE PROCESAMIENTO DE INFORMACION</b>		REG_EST_NACIONAL_2018								
<b>UBICACIÓN ARCHIVO DE PROCESAMIENTO</b>		<a href="#">\\WBSYSTEM11\REGADNAL</a>								
<b>TIPO DE ARCHIVO DE PROCESAMIENTO</b>		Sql Server								
<b>CANTIDAD DE VARIABLES</b>		57								
<b>CANTIDAD DE REGISTROS</b>		62.939.194								

Nº	NOMBRE VARIABLE	PRESENCIA DE DATOS		VALIDACION REGISTROS CON DATOS		INFORMACION DUPLICADA POR VARIABLE	
		REGISTROS VACIOS	REGISTROS CON DATOS	REGISTROS INCONSISTENTES	REGISTROS CONSISTENTES	DUPLICADOS DOCUMENTO	DUPLICADOS NOM Y F NAC
1	ID_TIPO_IDENTIFICACION	-	62.939.194	-	62.939.194	1.312	41.210
2	NUMERO_IDENTIFICACION	-	62.939.194	215	62.938.979	1.312	41.210
3	PRIMER_APELLIDO	129	62.939.065	12.717	62.926.348	1.312	41.210
4	SEGUNDO_APELLIDO	2.835.397	60.103.797	13.007	60.090.790	1.312	41.210
5	PRIMER_NOMBRE	624	62.938.570	9.352	62.929.218	1.312	41.210
6	SEGUNDO_NOMBRE	16.982.054	45.957.140	14.249	45.942.891	1.312	41.210
7	FECHA_NACIMIENTO	1.591.251	61.347.943	779.611	60.568.332	1.312	41.210
8	ID_SEXO	37.998	62.901.196	1	62.901.195	1.312	41.210

**VERIFICACION DE COMPLETITUD** (Diligencie la cantidad de registros según la completitud de la información válida, en los espacios en blanco, recuerde que la suma de los totales debe ser el total de registros)

**VERIFICACION DE NUMERO DE PERSONAS**

**OBSERVACIONES**

IDENTIFICACION DEL REGISTRO	DETALLE OBSERVACION

## Anexo 4. Inventario de registros administrativos

Nombre completo de la Entidad	Nombre del Registro Administrativo tal y como es solicitado	Variables Solicitadas	Periodicidad	Mecanismo de Solicitud	Datos del Contacto	Observaciones	Ruta Solicitud
Ministerio de Salud y Protección Social	* Base de Datos Unica de Afiliados BDUA * Base de Datos Unica de Afiliados régimen de excepción EDEX	tps_idn_id hst_idn_numero_identificacion afil_primer_apellido afil_segundo_apellido afil_primer_nombre afil_segundo_nombre afil_fecha_nacimiento tps_gnr_id dpr_id mnc_id ent_id ent_nombre afic_fecha_afiliacion_entidad tps_est_afil_id TIPO_ID_COTIZANTE NUMERO_ID_COTIZANTE TE tipo_cotizante tps_prn_id tps_cnd_bnf_id tps_rgm_id	Semestral	Correo Electrónico	HISNARDO ALBERTO UBAQUE UBAQUE Hisnardo.Ubaque@adres.gov.co CARLOS MARIO RAMIREZ RAMIREZ Carlos.Ramirez@adres.gov.co Dolly Esperanza Ovalle Carranza dovalle@minsalud.gov.co Diana Lopez dlopez@minsalud.gov.co Jeannett Adriana Umaná G jumana@minsalud.gov.co	octubre 2018	\\bherediag\Documentos\Solicitudes\BDUA Ministerio Corte Dic 2017 Nacional.docx \\mrescobarm\Diagnosticos\RA_2017\99_Cartas\Solicitudes\Censos1_ADRES_BDUASolicitud.tif