

	<b>PROCEDIMIENTO CALIDAD DE DATOS</b>	Código: DE04-P01
		Versión: 1
		Página 1 de 17

## CONTENIDO

1. OBJETIVO.....	3
2. DESTINATARIOS.....	3
3. GLOSARIO.....	3
4. REFERENCIAS.....	5
5. GENERALIDADES.....	5
6. REPRESENTACIÓN ESQUEMÁTICA DEL PROCEDIMIENTO.....	10
7. DESCRIPCIÓN DE ETAPAS Y ACTIVIDADES.....	12
7.1. COMPRENDER EL REQUERIMIENTO.....	12
7.1.1. Identificar los hallazgos de calidad de datos.....	12
7.1.2. Identificar fuentes de datos y procesos de transformación de datos...	12
7.2. VALIDAR Y PERFILAR LOS DATOS.....	12
7.2.1. Perfilar los datos y generar informe con estado inicial.....	12
7.2.2. Validar uso de datos maestros y de referencia.....	13
7.2.3. Establecer condiciones esperadas.....	13
7.2.4. Reportar hallazgos.....	13
7.3. ESTABLECER REGLAS DE NEGOCIO.....	14
7.3.1. Definir las reglas de negocio.....	14
7.3.2. Diseñar la solución de calidad de datos.....	15
7.4. CORREGIR, LIMPIAR Y MEJORAR LOS DATOS.....	16
7.4.1. Implementar la solución de calidad de datos.....	16
7.4.2. Realizar pruebas y validar mejora de calidad de los datos.....	16
7.5. DOCUMENTAR Y NOTIFICAR EL PROCESO DE CALIDAD DE DATOS	17
7.5.1. Documentar cada etapa del proceso ejecutada.....	17
7.5.2. Notificar el resultado del proceso.....	17
8. DOCUMENTOS RELACIONADOS.....	17

Elaborado por:  Nombre: Rodrigo Andrés Bocanegra  Cargo: Contratista Oficina de Tecnología e Informática	Revisado y Aprobado por:  Nombre: Francisco Andrés Rodríguez Erazo  Cargo: Jefe Oficina de Tecnología e Informática	Aprobación Metodológica por:  Nombre: Giselle Johanna Castelblanco Muñoz  Cargo: Representante de la Dirección para el Sistema de Gestión de Calidad  Fecha: 2020-10-02
--	---	---

Cualquier copia impresa, electrónica o de reproducción de este documento sin la marca de agua o el sello de control de documentos, se constituye en copia no controlada.

	<b>PROCEDIMIENTO CALIDAD DE DATOS</b>	Código: DE04-P01
		Versión: 1
		Página 2 de 17

9. RESUMEN CAMBIOS RESPECTO A LA ANTERIOR VERSIÓN ..... 17

COPIA NO CONTROLADA

	<b>PROCEDIMIENTO CALIDAD DE DATOS</b>	Código: DE04-P01
		Versión: 1
		Página 3 de 17

## 1. OBJETIVO

Establecer los lineamientos para definir, controlar y mejorar la calidad de los datos de la Superintendencia de Industria y Comercio, con el fin de mejorar la confiabilidad de los servicios de información provistos y minimizar los riesgos en la toma de decisiones por datos incorrectos o fuera de contexto.

## 2. DESTINATARIOS

Este documento debe ser conocido y aplicado por todos aquellos servidores públicos o contratistas que participen directa o indirectamente en el procedimiento Calidad de Datos derivado del proceso de Gestión Estratégica de Tecnologías de la Información.

## 3. GLOSARIO

**ARQUITECTURA DE DATOS E INFORMACIÓN:** Define la estructura con la cual está representada y almacenada la información de una organización, lo mismo que los servicios de información existentes y que soporta. Incluye el modelo conceptual, el modelo de indicadores, los componentes de información y sus relaciones, y la representación lógica y física de los datos, entre otros. Esta arquitectura expresa también la relación que tiene con la arquitectura misional y con las demás arquitecturas de TI.

**BODEGA DE DATOS □ DWH 2.0:** Una bodega de datos es un repositorio unificado de información institucional que integra datos provenientes de diferentes Fuentes a través de procesos de limpieza, depuración y estandarización de los datos, con el fin de disponer la información empresarial a los diferentes usuarios y apoyar la toma de decisiones de la SIC. Una bodega de datos de segunda generación o DWH 2.0 está en capacidad además de brindar servicios que apoyen la operación de los sistemas de información, así como manejar un alto volumen de datos, datos no estructurados y de alta velocidad en un entorno Big Data.

**CALIDAD DEL DATO:** Calidad de un conjunto de información recogida en una base de datos, un sistema de información o una bodega de datos, que reúne entre sus atributos la exactitud, completitud, integridad, actualización, coherencia, relevancia, accesibilidad y confiabilidad<sup>1</sup> necesarias para resultar útiles al procesamiento, análisis y cualquier otro fin que un usuario quiera darles.

**DATO:** Es una representación simbólica de una característica particular de un elemento o situación. Tiene un tipo (por ejemplo, numérico, cadena de caracteres o

<sup>1</sup> Estos atributos se definen en **Dimensiones de calidad** de la sección **5. GENERALIDADES**.

	<b>PROCEDIMIENTO CALIDAD DE DATOS</b>	Código: DE04-P01
		Versión: 1
		Página 4 de 17

lógico) que determina el conjunto de valores que el dato puede tomar. En el contexto informático, los datos se almacenan, procesan y comunican usando medios electrónicos. Constituyen los elementos primarios de los sistemas de información.

**DATOS DE REFERENCIA:** Datos no transaccionales que contienen un conjunto de valores predefinidos que se utilizan para caracterizar otros datos en la entidad y que se definen más allá del contexto de la Entidad. Estos datos deben tener idealmente la misma definición en todos los sistemas de información de la entidad.

Algunos ejemplos de datos de referencia son: países con código ISO (Colombia, COL); el tipo de documento de identificación nacional, el sexo, el género, entre otros (definidos por la Registraduría Nacional u otras entidades).

**DATOS MAESTROS:** Datos transaccionales y no transaccionales que deben ser idealmente compartidos por todos los sistemas de información de la Entidad para mantener la consistencia institucional en los análisis. Estos datos son relevantes dentro del contexto de la Entidad y no fuera de ella.

Ejemplos de datos maestros no transaccionales pueden ser las dependencias y sus códigos, los procesos, los trámites y servicios de la SIC, entre otros. Ejemplos de datos maestros transaccionales pueden ser los datos de funcionarios, ciudadanos atendidos por la SIC, sanciones, actos administrativos, entre otros.

**DATOS TRANSACCIONALES:** Datos que se generan durante una transacción y que no tienen un conjunto de datos posibles predefinidos. Ejemplos de estos datos son los nombres de una persona, teléfono, solicitud, entre otros.

**ETL:** Extract, Transform and Load («extraer, transformar y cargar», frecuentemente abreviado ETL) es el proceso que permite a las organizaciones mover datos desde múltiples fuentes, reformatearlos y limpiarlos, y cargarlos en otra base de datos, o bodega de datos para analizar.

**INFORMACIÓN:** Es un conjunto de datos organizados y procesados que tienen un significado, relevancia, propósito y contexto. La información sirve como evidencia de las actuaciones de las entidades. Un documento se considera información y debe ser gestionado como tal.

**SIC:** Superintendencia de Industria y Comercio

	<b>PROCEDIMIENTO CALIDAD DE DATOS</b>	Código: DE04-P01
		Versión: 1
		Página 5 de 17

#### 4. REFERENCIAS

Jerarquía de la norma	Numero/ Fecha	Título	Artículo	Aplicación Específica
Decreto	1008 de 2018	Por el cual se establecen los lineamientos generales de la política de Gobierno Digital y se subroga el capítulo 1 del título 9 de la parte 2 del libro 2 del Decreto 1078 de 2015, Decreto Único Reglamentario del sector de Tecnologías de la Información y las Comunicaciones.	Sección 2	Brinda lineamientos de arquitectura de datos e información conforme al marco de arquitectura TI Colombia.
CONPES	3920 de 2018	Política Nacional de Explotación de Datos (Big Data).	5	Brinda lineamientos específicos para masificar la explotación de datos de las entidades públicas.

#### 5. GENERALIDADES

La falta de calidad de los datos es uno de los principales problemas a los que se enfrentan los responsables de sistemas de información y los usuarios de los sistemas que requieren analizar la operación o tomar decisiones estratégicas basada en datos. Las funciones de TI para homologar y combinar fuentes de datos, crear copias de datos, poblar retrospectivamente, y especialmente los mecanismos de integración e interoperabilidad de los datos entre diferentes sistemas de información, generan riesgos de calidad de datos que deben ser gestionados.

La calidad de los datos es sinónimo de calidad de la información ya que los malos resultados de calidad de datos conllevan a información inexacta y bajo rendimiento del negocio. La limpieza de datos puede dar lugar a mejoras costosas de corto plazo que no abordan las causas raíz de los defectos de datos. Un programa más riguroso de calidad de datos es necesario para proporcionar una solución más económica para la mejora de la calidad y la integridad de los datos.

La SIC al gestionar la calidad de sus datos obtiene los siguientes beneficios:

- Minimizar los riesgos en los proyectos de la Entidad que requieren insumos de datos y en la toma de decisiones.

	<b>PROCEDIMIENTO CALIDAD DE DATOS</b>	Código: DE04-P01
		Versión: 1
		Página 6 de 17

- ▯ Ahorro de tiempo y recursos, haciendo un mejor uso de los recursos tecnológicos y humanos.
- ▯ Toma de decisiones oportunas con información confiable y de calidad.
- ▯ Adaptación a estándares o regulaciones internacionales sobre el manejo de información, permitiendo fortalecer la calidad de datos de la Entidad.
- ▯ Mejorar la confianza en los procesos, los sistemas de información y en los datos subyacentes.

### **DIMENSIONES DE CALIDAD DE LOS DATOS**

Las dimensiones y/ o criterios de la calidad de los datos para la SIC, incluyen:

- ▯ **Exactitud:** La exactitud de datos se refiere a la medida en que los datos representan correctamente las entidades de la "la vida real" que estos modelan. En muchos casos, la exactitud se mide por cómo los valores coinciden con una fuente de referencia identificada de información correcta.
- ▯ **Completitud:** Una de las expectativas acerca de la completitud de los datos es que ciertos atributos siempre tengan valores asignados en un conjunto de datos. La completitud puede comprenderse como la usabilidad y la adecuación de los valores de datos.
- ▯ **Consistencia:** La consistencia se refiere a garantizar que los valores de datos en un conjunto de datos son coherentes con los valores en otro conjunto de datos. El concepto de consistencia es relativamente amplio; puede incluir una expectativa de que dos valores de datos extraídos de dos conjuntos de datos separados no deben entrar en conflicto entre sí, o definir la coherencia con un conjunto de restricciones preestablecidas.
- ▯ **Vigencia:** La vigencia de los datos se refiere al grado en el que la información está al día con el mundo que modela. La vigencia mide qué tan actualizados están los datos, como también qué tan correctos están frente a los posibles cambios relacionados con el tiempo. Reglas de vigencia de los datos definen el "ciclo de vida" de un valor de datos antes de que caduque o de que requiera una actualización.
- ▯ **Precisión:** La precisión se refiere al nivel de detalle del elemento de datos. Los datos numéricos pueden necesitar precisión a varios dígitos significativos. Por

	<b>PROCEDIMIENTO CALIDAD DE DATOS</b>	Código: DE04-P01
		Versión: 1
		Página 7 de 17

ejemplo, el redondeo y truncamiento puede dar lugar a errores en que es necesaria una precisión exacta.

- ▯ Privacidad: Privacidad se refiere a la necesidad de control de acceso y de supervisión del uso. Algunos elementos de datos requieren restricción de uso o acceso.
- ▯ Razonabilidad: La razonabilidad debe considerarse cuando las expectativas de consistencia son relevantes dentro de contextos operacionales específicos. Por ejemplo, se podría esperar que el número de transacciones cada día no supere el 105% del número promedio de ejecución de las transacciones de los últimos 30 días.
- ▯ Integridad Referencial: La integridad referencial es la condición que se da cuando todas las referencias o relaciones destinadas a partir los de datos de una columna de una tabla hacia los datos de otra columna de esa o de diferente tabla son válidas. Las reglas de integridad referencial también se manifiestan como restricciones de duplicidad, para asegurar que cada entidad se produce una vez y sólo una vez.
- ▯ Oportunidad: La oportunidad se refiere a la expectativa de accesibilidad y disponibilidad de la información. Como ejemplo, medir un aspecto de la oportunidad como el tiempo transcurrido entre el punto en el que se espera la información y el punto en el cual está disponible para su uso.
- ▯ Unicidad: En esencia, la unicidad hace referencia a que ninguna entidad de datos exista más de una vez dentro de un conjunto de datos. Asegurar la unicidad de las entidades dentro un conjunto de datos implica que ninguna entidad existe más de una vez y que un valor clave referido a cada entidad es único y sólo es referido a esa entidad específica, dentro del conjunto de datos. La unicidad también se puede alcanzar cuando la redundancia en los datos está monitoreada y controlada, normalmente mediante técnicas de sincronización y orquestación como se realiza con datos maestros y de referencia.
- ▯ Validez: La validez se refiere a si las instancias de datos se almacenan, intercambian, o si se presentan en un formato que es consistente con el dominio de los valores, así como la consistencia con otros valores de atributos similares. La validez asegura que los valores de datos se ajustan a numerosos atributos asociados con el elemento de datos: su tipo de datos, precisión, patrones de formato, el uso de una enumeración de valores predefinidos, rangos de dominio, formatos de almacenamiento, entre otros.

	<b>PROCEDIMIENTO CALIDAD DE DATOS</b>	Código: DE04-P01
		Versión: 1
		Página 8 de 17

## ROLES INVOLUCRADOS

Los roles en este procedimiento se definieron según los lineamientos del Marco de Arquitectura Empresarial del Ministerio de Tecnología.

- ▮ **CIO:** Corresponde al jefe de la Oficina de Tecnología e Informática quien coordina los procesos de Tecnología e Información.
- ▮ **Gestor de Información:** Asegura con el CIO el cumplimiento de los planes, políticas y lineamientos definidos para la gestión de datos y la información. Administra los requerimientos de datos de la SIC y articula las personas, procesos y herramientas requeridos para atenderlos.
- ▮ **Analista de negocio:** Analista funcional que conoce en detalle las actividades y métricas de los procesos de negocio, es decir los procesos misionales, estratégicos, evaluación y de apoyo, que serán impactados por proyectos de gestión de datos e información. Define la prioridad en los servicios de información a implementar, las reglas de negocio para mejorar la calidad de los datos y asegura que los servicios de información entregados por la OTI sean confiables y útiles para los demás interesados en el proceso.

Todos los líderes de proceso se consideran analistas de negocio, quienes son propietarios y responsables de los datos que se generan en el desarrollo de su proceso. Adicionalmente, los líderes de proceso podrán designar otros analistas de negocio que apoyan en la gestión con el personal de la Oficina de Tecnología e Informática las necesidades de datos e información del proceso. Puede adicionalmente ejercer el rol de Publicador.

- ▮ **Analista de información:** Encargado de estructurar la información y los datos para apoyar la definición de los requerimientos y alinearlos con la Arquitectura de datos institucional.
- ▮ **Ingeniero de calidad de datos:** Especializado en actividades de perfilamiento y calidad de datos. Este rol lo puede desempeñar un ingeniero de software o un ingeniero de datos, dado que los requerimientos de calidad de datos se pueden gestionar a nivel de sistemas de información (registro de datos en formularios, validaciones del sistema, procedimientos calculados, bases de datos transaccionales) o a nivel de capa de datos (bodega de datos, bases de datos analíticas, datos maestros o de referencia, entre otros).



 Industria y Comercio SUPERINTENDENCIA	PROCEDIMIENTO CALIDAD DE DATOS	Código: DE04-P01
		Versión: 1
		Página 9 de 17

## SISTEMAS DE INFORMACIÓN Y OTROS MEDIOS ELECTRÓNICOS

### Sistemas Internos

- ▮ Repositorio Gestión de Información ▮ Carpeta ▮10. Calidad de datos▮  
<https://drive.google.com/drive/u/1/folders/0AGM4MZIUcBhmUk9PVA>
- ▮ Gestión de proyecto  
<https://dev.azure.com/SUPER-INDUSTRIA/SIC%20ANALITICA/>
- ▮ Procesamiento de datos de La Superintendencia de Industria y Comercio  
<https://portal.azure.com/> - RG-DataWarehouse
- ▮ Sistema Integral de Gestión Institucional  
<https://sigi.sic.gov.co/SIGI/portal/index.php>

COPIA NO CONTROLADA

## 6. REPRESENTACIÓN ESQUEMÁTICA DEL PROCEDIMIENTO

No	ETAPAS	ENTRADAS	DESCRIPCIÓN DE LA ETAPA	RESPONSABLE	SALIDAS
1	<b>COMPRENDER EL REQUERIMIENTO</b>	Requerimientos de datos	<p>De acuerdo con los requerimientos de datos identificados en la gestión de datos e información. establece requerimientos específicos de calidad de datos y detalla los insumos necesarios para atender el requerimiento.</p> <p>En esta etapa se desarrollan las siguientes actividades:</p> <ul style="list-style-type: none"> <li>- Identificar los hallazgos de calidad de datos</li> <li>- Identificar fuentes de datos y procesos de transformación de datos</li> </ul>	Gestor de Información	<p>DE04-F03 Formato Informe de calidad de datos: Requerimiento de calidad de datos</p>
2	<b>VALIDAR Y PERFILAR LOS DATOS</b>	<p>Fuentes de datos y bodegas de datos</p> <p>Metadatos</p> <p>Datos Maestros y de referencia</p> <p>Requerimiento de calidad de datos</p>	<p>Con base en el requerimiento de calidad de datos, realiza un perfilamiento de los datos, es decir, una muestra que permita comprender como se están comportando los datos, establece dependencias y relaciones con las fuentes de datos existentes, datos maestros y datos de referencia. Reporta los hallazgos según las condiciones esperadas de calidad.</p> <p>En esta etapa se desarrollan las siguientes actividades:</p> <ul style="list-style-type: none"> <li>- Perfilar los datos y generar informe con estado inicial</li> <li>- Validar uso de datos maestros y de referencia</li> <li>- Establecer condiciones esperadas</li> <li>- Reportar los hallazgos</li> </ul>	Gestor de Información	<p>DE04-F03 Formato Informe de calidad de datos:</p> <ul style="list-style-type: none"> <li>- Perfilamiento</li> <li>- Datos maestros o referencia aplicables</li> <li>- Condiciones esperadas</li> <li>- Hallazgos</li> </ul>

No	ETAPAS	ENTRADAS	DESCRIPCIÓN DE LA ETAPA	RESPONSABLE	SALIDAS
3	<b>ESTABLECER REGLAS DE NEGOCIO</b>	Perfilamiento Condiciones esperadas Hallazgos	<p>Con los datos perfilados, valida y diseña con el analista de negocio las reglas de negocio que se deben aplicar para corregir, limpiar y mejorar los datos.</p> <p>En esta etapa se desarrollan las siguientes actividades:</p> <ul style="list-style-type: none"> <li>- Definir las reglas de negocio</li> <li>- Diseñar la solución de calidad de datos</li> </ul>	Ingeniero de calidad de datos	<p>DE04-F03 Formato Informe de calidad de datos:</p> <ul style="list-style-type: none"> <li>- Reglas de negocio</li> <li>- Diseño de la solución</li> </ul>
4	<b>CORREGIR, LIMPIAR Y MEJORAR LOS DATOS</b>	Fuentes de datos y bodegas de datos Reglas de negocio Diseño de la solución	<p>Realiza las actividades planteadas para corregir, limpiar y mejorar los datos. Las actividades se desarrollan sobre las fuentes de información que hayan sido acordadas con el analista de negocio, procurando siempre que se ejecuten desde las fuentes primarias de información.</p> <p>En esta etapa se desarrollan las siguientes actividades:</p> <ul style="list-style-type: none"> <li>- Implementar la solución de calidad de datos</li> <li>- Realizar pruebas y validar mejora de calidad de los datos</li> </ul>	Ingeniero de calidad de datos	<p>Código fuente</p> <p>Datos limpiados, corregidos y mejorados</p> <p>DE04-F03 Formato Informe de calidad de datos: Resultados</p>
5	<b>DOCUMENTAR Y NOTIFICAR EL PROCESO DE CALIDAD DE DATOS</b>	Datos limpiados, corregidos y mejorados	<p>Realiza una revisión del proceso de calidad de datos aplicado y el impacto generado a los usuarios finales. Involucra los resultados del proceso de calidad de datos dentro de la gestión de datos e información y notifica el resultado.</p> <p>En esta etapa se desarrollan las siguientes actividades:</p> <ul style="list-style-type: none"> <li>- Documentar cada etapa del proceso ejecutada</li> <li>- Notificar el resultado del proceso</li> </ul>	Ingeniero de calidad de datos	<p>DE04-F03 Formato Informe de calidad de datos</p> <p>Metadatos reglas de negocio</p> <p>Notificación del proceso de calidad de datos (acta, correo)</p>

	<b>PROCEDIMIENTO CALIDAD DE DATOS</b>	Código: DE04-P01
		Versión: 1
		Página 12 de 17

## **7. DESCRIPCION DE ETAPAS Y ACTIVIDADES**

### **7.1. COMPRENDER EL REQUERIMIENTO**

#### **7.1.1. Identificar los hallazgos de calidad de datos**

El analista de negocio reporta los hallazgos de calidad de datos sobre servicios de información que tenga en uso, según se describe en la actividad 7.1.1 del DE04-P02 Procedimiento Gestión de Datos e Información.

El gestor de información define requerimientos dentro de un proceso de integración de datos en los que se haya identificado gran cantidad de problemas de calidad dentro de un ejercicio de arquitectura empresarial o diagnóstico técnico.

El analista de datos detalla la definición del requerimiento de calidad de datos y lo documenta en la herramienta de gestión de proyectos y desarrollo dispuesta para tal fin.

#### **7.1.2. Identificar fuentes de datos y procesos de transformación de datos**

A través de los datos identificados y reportados por los usuarios finales, el analista de información realiza la traza e identificación de las fuentes de información y todas las capas de tratamiento de los datos para establecer las actividades concretas de corrección por cada capa (origen, extracción, consolidación, transformación con reglas de negocio, análisis).

En esta identificación se complementa y detalla el requerimiento de calidad de datos, realizando adicionalmente la priorización del requerimiento respecto al portafolio de requerimientos de datos según se describe en la etapa 7.1 del DE04-P02 procedimiento Gestión de Datos e Información.

### **7.2. VALIDAR Y PERFILAR LOS DATOS**

#### **7.2.1. Perfilar los datos y generar informe con estado inicial**

El analista de información analiza los datos con técnicas estadísticas y de perfilamiento para identificar los problemas de calidad, haciendo uso de herramientas especializadas para este tipo de labores.

El resultado del análisis se incluye en el Informe de calidad de datos, sección perfilamiento, el cual puede incluir recursos y documentos adicionales generados por la herramienta de perfilamiento. Estos recursos se vinculan al requerimiento de calidad de datos en la herramienta de gestión.

	<b>PROCEDIMIENTO CALIDAD DE DATOS</b>	Código: DE04-P01
		Versión: 1
		Página 13 de 17

### **7.2.2. Validar uso de datos maestros y de referencia**

El analista de información por cada requerimiento de calidad de datos identificado realiza la validación del uso de fuentes de datos, datos maestros y datos de referencia, con el fin de establecer actividades que permitan alinearse con la arquitectura de datos institucional.

Esta validación permite orientar la definición de reglas de negocio y el diseño de la solución en etapas posteriores.

El resultado de esta actividad se documenta en el Formato Informe de calidad de datos, sección: Datos maestros o referencia aplicables; y se vincula al requerimiento de calidad de datos en la herramienta de gestión.

### **7.2.3. Establecer condiciones esperadas**

El analista de información con el analista de negocio, definen las condiciones que se esperan cumplir para cada uno de los requerimientos de datos identificados y definen el umbral máximo permitido de error sobre la cantidad total de registros.

Si bien el ideal es manejar un umbral de 0% (0 registros erróneos / cantidad total de registros), se debe tener en cuenta que en algunos casos la información de validación puede provenir de otras fuentes con problemas de calidad que no pueden resueltos por la SIC (por ejemplo: la lista de cámaras de comercio que proviene de otra entidad: Confecámaras). En estos casos el umbral se definirá según la cantidad de registros mínimos que permiten operar con una disminución significativa de reprocesos o cuando sea posible realizar un análisis estadístico de los datos.

Las condiciones esperadas de la calidad de datos se registran en el DE04-F03 Formato Informe de calidad de datos. Estas condiciones pueden cambiar al realizar la validación de aplicación de reglas de negocio, según se describe en la actividad 7.4.2 Realizar pruebas y validar mejora de calidad de los datos.

El resultado de esta actividad se documenta en el Formato Informe de calidad de datos, sección: Condiciones esperadas; y se vincula al requerimiento de calidad de datos en la herramienta de gestión.

### **7.2.4. Reportar hallazgos**

De acuerdo con el perfilamiento de los datos y las condiciones esperadas, el analista de información realiza un reporte de los hallazgos al analista de negocio en el DE04-F03 Formato Informe de calidad de datos. Allí se relacionan las variables fuera de los valores esperados y la comparación con mallas de validación generales de datos que hacen parte de la documentación de los metadatos.

	<b>PROCEDIMIENTO CALIDAD DE DATOS</b>	Código: DE04-P01
		Versión: 1
		Página 14 de 17

Los hallazgos relacionan la cantidad de registros que se verifican, los requisitos de calidad y la cantidad de registros que no cumplen la validación, con el fin de comparar la brecha entre la calidad esperada y la calidad actual de los datos.

El resultado de esta actividad se documenta en el Formato Informe de calidad de datos, sección: Hallazgos; y se vincula al requerimiento de calidad de datos en la herramienta de gestión.

### 7.3. ESTABLECER REGLAS DE NEGOCIO

#### 7.3.1. Definir las reglas de negocio

Con los datos perfilados, el analista de información se reúne con el analista de negocio y el ingeniero de calidad de datos para determinar las reglas de negocio que se deben aplicar con el fin de corregir, limpiar y mejorar los datos.

Se establece un alcance de los datos que se espera limpiar, corregir y/o mejorar en las bases de datos de los sistemas de información de la SIC.

- **Limpiar** implica que los datos se ajustan y estandarizan, pero sin modificar su significado.
  - **Ejemplo:** Cuando se encuentren las siguientes coincidencias □BOGOTÁ D.C.□, □BOGOTA□, □BOGOTÁ□, □BOGOTA DC□, quedarán estandarizados como □BOGOTÁ DC□.
- **Corregir** implica que los datos cambian su contenido y eventualmente su significado.
  - **Ejemplo:** Para las personas que tengan una edad superior a 100 años con base en el registro de fecha de nacimiento del formulario de inscripción, recalcula la edad con base en la fecha de nacimiento indicada en el registro civil de nacimiento de la Registraduría Nacional.
- **Mejorar** implica que los datos son complementados con datos adicionales que brindan información adicional al usuario final.
  - **Ejemplo:** Las actuaciones registradas como □DENUNCIA 0□ y □TRASLADO□ se agruparán en un nuevo concepto denominado **etapa** y definido como □PRESENTACIÓN□. Normalmente se define a través de tablas con la relación de los datos y las nuevas columnas incluidas.

Estas definiciones se realizan con base en reglas de negocio o según nuevas taxonomías (árboles conceptuales) identificadas con el analista de negocio.

	<b>PROCEDIMIENTO CALIDAD DE DATOS</b>	Código: DE04-P01
		Versión: 1
		Página 15 de 17

El resultado de esta actividad se documenta en el Formato Informe de calidad de datos, sección: Reglas de negocio; y se vincula al requerimiento de calidad de datos en la herramienta de gestión.

### 7.3.2. Diseñar la solución de calidad de datos

El ingeniero de calidad de datos designado por el líder del sistema de información o el gestor de información, documenta el diseño de la solución teniendo en cuenta el análisis previo de los datos, el perfilamiento, las reglas de negocio, los lineamientos y los estándares. Adicionalmente debe tener en cuenta los lineamientos de calidad de datos, en los cuales establece que se debe procurar ajustar la calidad de datos desde la fuente del dato, es decir, desde el sistema o servicio que crea la información.

El diseño de la solución puede considerar algunos de los siguientes tipos de mejora y se registra en el DE04-F03 Formato Informe de calidad de datos:

#### Sistema de información

- ▢ Validación en el ingreso de datos a través de formularios electrónicos
- ▢ Validación de datos en las transacciones de negocio del sistema de información
- ▢ Corrección de datos históricos en la base de datos transaccional, considerando los efectos que esta corrección pueda tener en el funcionamiento del sistema

#### Capa de datos

- ▢ Transformación de datos en la capa de negocio de la bodega de datos (denominado Business Vault). Nunca se realizará transformación en la capa de datos crudos (datos que provienen directamente del sistema de información y denominados Raw Vault).
- ▢ Ajustes en las métricas, agrupaciones o cálculos en los modelos analíticos de informes.
- ▢ Filtros y segmentaciones específicos en informes con indicadores.

El diseño de la solución en la capa de datos se realiza según lo establece el documento **Atención de requerimientos de bodega de datos e informes analíticos.**

El resultado de esta actividad se documenta en el Formato Informe de calidad de datos, sección: Diseño de la solución; y se vincula al requerimiento de calidad de datos en la herramienta de gestión.

	<b>PROCEDIMIENTO CALIDAD DE DATOS</b>	Código: DE04-P01
		Versión: 1
		Página 16 de 17

## 7.4. CORREGIR, LIMPIAR Y MEJORAR LOS DATOS

### 7.4.1. Implementar la solución de calidad de datos

El ingeniero de calidad de datos designado implementa técnicamente las actividades definidas en la fase de diseño, ajustando los datos en el destino de la información seleccionado: bodega de datos o sistema de información.

La implementación de la solución se realiza de acuerdo con los procedimientos:

- ▯ GS03-P03 Procedimiento Ciclo de vida de Construcción de Software: Para los ajustes realizados sobre sistemas de información.
- ▯ DE04-P02 Procedimiento Gestión de datos e información: Para los ajustes realizados sobre la capa de datos: bodega de datos, datos maestros y referencia, bases de datos analíticas.

En los cuales se establece la generación del código fuente y cuyo resultado de implementación deben ser los nuevos datos limpios, corregidos y mejorados.

### 7.4.2. Realizar pruebas y validar mejora de calidad de los datos

El ingeniero de calidad de datos contrasta los datos antes y después del proceso de corrección para validar la efectividad de las acciones aplicadas, con el fin de establecer si se logró cumplir con la calidad esperada definida previamente.

El analista de información presenta los resultados al analista de negocio delegado por el director o jefe de área, para que se apruebe el proceso de calidad de datos realizado en el DE04-F03 Formato Informe de calidad de datos.

Si cumple con la calidad esperada, genera los registros de salida y continua con la actividad 7.5.1 Documentar cada etapa del proceso ejecutada.

De lo contrario, reevalúa el ejercicio y regresa a la actividad 7.2.1 Perfilar los datos y generar informe con estado inicial.

El resultado de esta actividad se documenta en el Formato Informe de calidad de datos, sección: Resultados; y se vincula al requerimiento de calidad de datos en la herramienta de gestión.



	<b>PROCEDIMIENTO CALIDAD DE DATOS</b>	Código: DE04-P01
		Versión: 1
		Página 17 de 17

## 7.5. DOCUMENTAR Y NOTIFICAR EL PROCESO DE CALIDAD DE DATOS

### 7.5.1. Documentar cada etapa del proceso ejecutada

El ingeniero de calidad de datos valida y consolida toda la documentación en el informe de calidad de datos y asegura el registro de los metadatos aplicables a la base de datos, bodega de datos, o demás bases utilizadas que cuenten con registro de metadatos.

El resultado de esta actividad se verifica la documentación de todas las secciones del Formato Informe de calidad de datos; y se vincula al requerimiento de calidad de datos en la herramienta de gestión.

### 7.5.2. Notificar el resultado del proceso

El gestor de información notifica al líder de proceso y al analista de negocio el resultado del proceso realizado según el requerimiento acordado, adjuntando la versión final y revisada del DE04-F03 Formato Informe de calidad de datos. El registro de la notificación quedará mediante acta o correo electrónico.

Continúa con la actividad 7.5.2 del procedimiento Gestión de Datos e Información DE04-P02.

## 8. DOCUMENTOS RELACIONADOS

SC05-I01 Políticas del Sistema de Gestión de Seguridad de la Información □ SGSI  
 Política de Gestión de Datos e Información  
 GS03-P03 Procedimiento Ciclo de vida de Construcción de Software  
 DE04-P02 Procedimiento Gestión de datos e información  
 DE04-F03 Formato Informe de calidad de datos  
 Atención de requerimientos de bodega de datos e informes analíticos  
 Marco de arquitectura empresarial MinTIC: Dominio de Información  
 Políticas y lineamientos de Gobierno Digital

## 9. RESUMEN CAMBIOS RESPECTO A LA ANTERIOR VERSIÓN

Creación del procedimiento.
-----------------------------