

## GESTIÓN DEL CONOCIMIENTO

### Instrucciones para la Calidad de los Datos

Código: GCN-IN006

Versión: 01

Vigencia desde:  
30 de marzo de 2023

#### Control de cambios

Versión	Fecha	Descripción de la modificación
01	30 de marzo de 2023	Creación de las Instrucciones para la calidad de los datos.

Método de Elaboración	Revisa	Aprueba
Documento elaborado por el Equipo de Analítica Institucional con el acompañamiento metodológico de la analista de la Oficina Asesora de Planeación – Analista de proceso	<p style="text-align: center;"><b>Yolima Elisabeth López Forero</b> Profesional OAP – Equipo Analítica Institucional</p> <p style="text-align: center;"><b>Yamile Espinosa Galindo</b> Profesional OAP – Analista del proceso</p>	<p style="text-align: center;"><b>Gabriel Felipe Angarita Serrano</b> Jefe Oficina Asesora de Planeación Líder del proceso Gestión del Conocimiento</p> <p style="text-align: center;">Caso Hola No. <b>312622</b></p>

***Nota:** Por responsabilidad ambiental no imprima este documento. Si este documento se encuentra impreso se considera “Copia no Controlada”. La versión vigente se encuentra publicada en la intranet de la Secretaría Distrital de Gobierno”*

## 1. INFORMACIÓN GENERAL

### Propósito

Establecer lineamientos para la captura, validación, coherencia y calidad de los datos que se producen y/o cargan en las herramientas de procesamiento y sistemas de información de la Secretaría Distrital de Gobierno, con el propósito de ser útil y producir valor para los diferentes públicos para la toma de decisiones estratégicas en la entidad.

### Responsable

Jefe(a) Oficina Asesora de Planeación

### Glosario

**Accesibilidad:** Grado en el que los datos pueden ser accedidos en un contexto específico, particularmente por personas que necesiten tecnologías de apoyo o una configuración especial por algún tipo de discapacidad.

**Actualidad:** Grado en el que los datos tienen atributos que tienen la edad correcta en un contexto de uso específico.

**Calidad de datos:** Es el ámbito enfocado en el aseguramiento de la calidad para garantizar la prestación de servicios de información e institucionales, a través de la identificación y propuesta de mejoras, la modificación del modelo operativo y la actualización y verificación del cumplimiento de los indicadores de calidad definidos para el dato.

**Ciclo de vida de datos:** Es el ámbito enfocado en asegurar el cubrimiento del enmascaramiento, reducción y archivado de los datos en la institución, soportando el nivel de autoservicio comprometido.

**Confidencialidad:** Grado en el que los datos tienen atributos que aseguran que los datos son sólo accedidos e interpretados por usuarios autorizados en un contexto de uso específico. La confidencialidad es un aspecto de la seguridad de la información (junto con la disponibilidad y la integridad) definida como en ISO/IEC 13335-1:2004.

**Conformidad:** Grado en el que los datos tienen atributos que se adhieren a estándares, convenciones o normativas vigentes y reglas similares referentes a la calidad de datos en un contexto de uso específico.

**Compleitud:** Grado en el que los datos asociados con una entidad tienen valores para todos los atributos esperados e instancias de entidades relacionadas en un contexto de uso específico.

**Comprensibilidad:** Grado en el que los datos tienen atributos que permiten ser leídos e interpretados por los usuarios y son expresados utilizando lenguajes, símbolos y unidades apropiados en un contexto de uso

específico. Cierta información sobre la comprensibilidad puede ser expresada mediante metadatos.

**Consistencia:** Grado en el que los datos están libres de contradicción y son coherentes con otros datos en un contexto de uso específico. Puede ser analizada en datos que se refieran tanto a una como a varias entidades comparables.

**Credibilidad:** Grado en el que los datos tienen atributos que se consideran ciertos y creíbles en un contexto de uso específico. La credibilidad incluye el concepto de autenticidad (la veracidad de los orígenes de datos, atribuciones, compromisos).

**Disponibilidad:** Grado en el que los datos tienen atributos que permiten ser obtenidos por usuarios y/o aplicaciones autorizados en un contexto de uso específico.

**Eficiencia:** Grado en el que los datos tienen atributos que pueden ser procesados y proporcionados con los niveles de rendimiento esperados mediante el uso de cantidades y tipos adecuados de recursos en un contexto de uso específico.

**Exactitud:** Grado en el que los datos representan correctamente el verdadero valor del atributo deseado de un concepto o evento en un contexto de uso específico. Tiene dos principales aspectos: Exactitud Sintáctica, Exactitud Semántica

**Exactitud Sintáctica:** cercanía de los valores de los datos a un conjunto de valores definidos en un dominio considerado sintácticamente correcto.

**Exactitud Semántica:** cercanía de los valores de los datos a un conjunto de valores definidos en un dominio considerado semánticamente correcto.

**Integridad:** se refiere a que todas las referencias o relaciones de la tabla a la que pertenece el dato (campo) son válidas en otros conjuntos de datos con los que se pueda relacionar.

**Oportunidad:** se refiere a la expectativa de accesibilidad y disponibilidad de la información. Están disponibles cuando se requiere.

**Portabilidad:** Grado en el que los datos tienen atributos que les permiten ser instalados, reemplazados o eliminados de un sistema a otro, preservando el nivel de calidad en un contexto de uso específico.

**Precisión:** Grado en el que los datos tienen atributos que son exactos o proporcionan discernimiento en un contexto de uso específico.

**Recuperabilidad:** Grado en el que los datos tienen atributos que permiten mantener y preservar un nivel específico de operaciones y calidad, incluso en caso de fallos, en un contexto de uso específico.

**Trazabilidad:** Grado en el que los datos tienen atributos que proporcionan un camino de acceso auditado a los datos o cualquier otro cambio realizado sobre los datos en un contexto de uso específico.

**Uso de información (valor en la transformación):** se refiere al uso de información de calidad en los procesos o iniciativas de migración / fusión de datos hacia una institución convergente, multicanal y con una oferta simplificada. Facilitando con esto la gestión de la transformación y el proceso de migración de datos.

**Validez:** el dato es válido si está conforme al tipo de dato esperado, rango o estándar según se haya definido. Se refiere a si las instancias de datos se almacenan, intercambian, o si se presentan en un formato que es válido para el dominio de los valores.

## Siglas

**DTI-** Dirección de Tecnologías de la Información

**ICONTEC-** Instituto Colombiano de Normas Técnicas y Certificación

**MINTIC-** Ministerio de Tecnologías de la Información y las Comunicaciones

**MIPG-** Modelo Integrado de Planeación y Gestión

**SDG-** Secretaría Distrital de Gobierno

## 2. INSTRUCCIONES

El presente documento de instrucciones desarrolla lo pertinente a la limpieza y validación de datos contenido en los siguientes procedimientos:

GCN-P006 Herramientas para uso y apropiación y analítica institucional

GCN-P009 Procedimiento gestión de información estadística

La calidad de los datos hace referencia al grado en que el producto de los datos satisface las necesidades y requerimientos definidos por la entidad, que impulsan la innovación, el desarrollo y la transformación. De acuerdo con lo establecido en la ISO/IEC 25012, las características de Calidad de Datos están clasificadas en dos grandes categorías: características inherentes y características dependientes del sistema.

- **Calidad de Datos Inherente:** Se refiere al grado con el que las características de calidad de los datos tienen el potencial intrínseco para satisfacer las necesidades establecidas y necesarias cuando los datos son utilizados bajo condiciones específicas. Desde el punto de vista inherente, la Calidad de Datos se refiere a los mismos datos, en particular a:
  - Valores de dominios de datos y posibles restricciones (e.g., Reglas de Negocio gobernando la calidad requerida por las características en una aplicación dada).
  - Relaciones entre valores de datos (e.g., Consistencia).
  - Metadatos.

Dentro de las características de calidad de datos inherentes, se encuentran:

- Exactitud
- Completitud
- Consistencia
- Credibilidad
- Actualidad

- **Calidad de Datos Dependiente del Sistema:** Se refiere al grado con el que la Calidad de Datos es alcanzada y preservada a través de un sistema informático cuando los datos son utilizados bajo condiciones específicas.

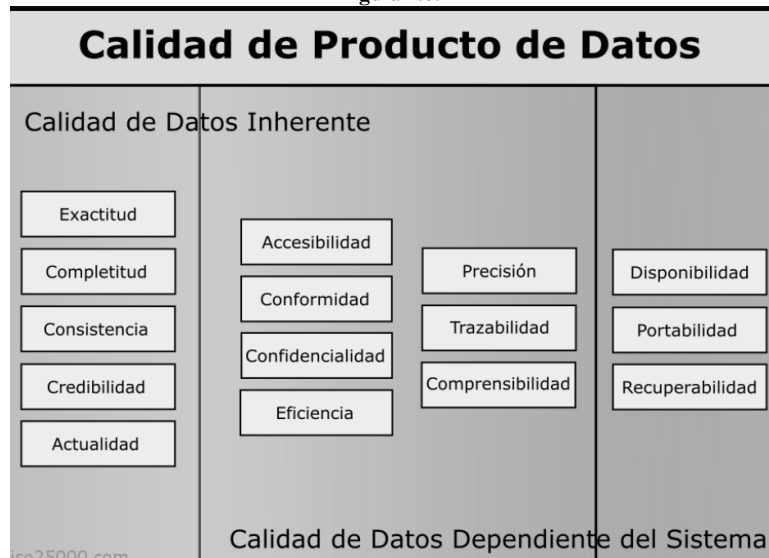
Desde el punto de vista dependiente del sistema, la Calidad de Datos depende del dominio tecnológico en el que los datos se utilizan, y se alcanza mediante las capacidades de los componentes del sistema informático tales como: dispositivos hardware (e.g., Respaldo Software para alcanzar la Recuperabilidad), y otro software (e.g., Herramientas de migración para alcanzar la Portabilidad).

Dentro de las características de calidad de datos dependiente del sistema, se encuentran:

- Accesibilidad
- Conformidad
- Confidencialidad
- Eficiencia
- Precisión
- Trazabilidad
- Comprensibilidad
- Disponibilidad
- Portabilidad
- Recuperabilidad

### Calidad de Producto de Datos

Figura No. 1



Fuente: ISO/IEC 25012,

A continuación se presentan los criterios que debe cumplir un conjunto de datos en la SDG, para que puedan ser de calidad e interoperables y cumplir su propósito fundamental de informar y analizar para poder tomar decisiones basadas en hechos y datos confiables, creíbles y seguros:

1. **Confidencialidad:** Este criterio establece que los datos solo deben ser accedidos por personas autorizadas para proteger adecuadamente la información reservada y clasificada. En Colombia, los criterios para asignar la calificación a las categorías de información son aquellos indicados por la Ley de Transparencia y del Derecho de Acceso a la Información (ley 1712 de 2014) la cual fue reglamentada con el Decreto 103 de 2015, en el que se establece que para la publicación de Datos Abiertos se debe tener en cuenta el título relativo a las “excepciones de acceso a la información”, el cual resalta la procedencia de limitar el acceso a la Información Pública Clasificada y la Información Pública Reservada.

El acceso a la información pública exige un ejercicio constante de ponderación entre el derecho a la información y el derecho a la protección de los datos personales. En este contexto surge la anonimización como una herramienta para mitigar los riesgos que presentan la obtención y el tratamiento masivo de los datos públicos, el cual es un proceso que permite identificar y ocultar la información sensible contenida en los documentos, permitiendo su divulgación sin que ello implique vulnerar los derechos a la protección de datos de las personas y entidades que se puedan referenciar en los mismos, garantizando que no sea posible identificar de manera directa o indirecta a individuos o entidades.

Una vez identificada la información clasificada y/o reservada, tenga en cuenta los siguientes elementos para anonimizar los datos a publicar en formatos abiertos:

- **Pre-anonimización:** Ningún conjunto de datos deberá permitir la identificación directa o indirecta de una persona. Por tal motivo, es recomendable eliminar la información personal con la que se va a trabajar y tener un especial cuidado con aquellos datos que contienen información sensible según la normatividad colombiana.
- **Ocultamiento, supresión y pseudo-anonimización:** Identifique en su conjunto de datos aquellos datos que sean sensibles para su eliminación o sustitución. A continuación, algunos ejemplos de datos sensibles:
  - Nombres.
  - Fecha de nacimiento y fecha de constitución en Cámara de Comercio, (excepto el año).
  - Números de teléfono y fax.
  - Números de identificación: cédula de ciudadanía, pasaporte y tarjeta de identidad. Números asociados a la seguridad social, licencias de conducción, Número de Identificación Tributaria (NIT), Registro Único Tributario (RUT), Registro Único de Proponentes (RUP) y Registro Único Empresarial (RUES).
  - Direcciones de correo electrónico.
  - Números de cuentas bancarias.
  - Identificadores del vehículo como placa, entre otros.
  - Identificadores de dispositivos móviles y números de serie.
  - Direcciones de IP.
  - Cualquier otro número único de identificación.

- Dirección de domicilio.

Una vez identificadas estas variables, se debe proceder a eliminar los campos del conjunto de datos o reemplazarlos por algún seudónimo.

- **Generalización:** se generalizan los atributos a través de la modificación de escalas o magnitudes (por ejemplo, cambiando una ciudad o municipio por un departamento, o días por meses). También puede ser aplicada creando rangos en lugar de valores específicos (por ejemplo, cambiar la edad absoluta por un rango de edades o grupo etario). Existen otras técnicas de anonimización descritas por el DANE para la producción de datos estadísticos que pueden ser consultadas en el siguiente enlace: [https://www.dane.gov.co/files/sen/lineamientos/DSO\\_020\\_LIN\\_08.pdf](https://www.dane.gov.co/files/sen/lineamientos/DSO_020_LIN_08.pdf)

2. **Relevancia:** Los datos publicados deben ser de utilidad para los usuarios. Determina los temas importantes, destacados, significativos o con mayor demanda por parte de los usuarios que se deben publicar en la apertura de datos. La SDG ha priorizado las temáticas de acuerdo con su misionalidad y funciones en los siguientes: <http://gaia.gobiernobogota.gov.co/content/dago>
3. **Actualidad:** Vigencia y actualidad de los datos publicados. La actualidad representa la vigencia y validez de los datos. Existen conjuntos de datos que se encuentran desactualizados en sus fechas, cifras periódicas o metadatos, por lo que es fundamental mantener la información actualizada constantemente. Aplique la prueba de actualidad a la última versión del conjunto de datos, pues es importante que permanezcan publicados los históricos de datos que son de gran importancia para la trazabilidad y los análisis temporales. Evalúe, además, que sus datos estén siendo publicados oportunamente en relación con la fecha en la que fueron producidos.

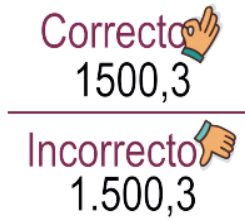
En el momento de publicar un conjunto de datos tenga en cuenta las siguientes consideraciones:

- Hay datos que pierden su vigencia y utilidad rápidamente, por eso establezca un límite de publicación más ajustado según las capacidades de la entidad.
  - Defina un cronograma de publicación según la periodicidad que se estipulan en los metadatos con el fin de que el conjunto de datos se conserve actualizado.
4. **Trazabilidad:** Histórico del conjunto de datos disponible: fechas de creación, publicación y actualizaciones. Este criterio permite determinar el flujo de procesos y acciones que se han generado sobre un conjunto de datos a través del tiempo. Este criterio es importante para la interoperabilidad de los conjuntos de datos entre las diferentes instituciones y para la reutilización de la información por parte de los usuarios. Garantice la trazabilidad publicando todos los históricos de los conjuntos de datos que se tengan disponibles. A manera de ejemplo, se recomienda consolidar todos los datos dispersos que están agrupados por vigencias o años en un solo conjunto de datos en donde se incluya una columna con el año de publicación de dichos datos.
  5. **Conformidad:** La conformidad hace referencia al cumplimiento de lineamientos y estándares

vigentes en el proceso de apertura de datos. En términos de calidad de datos, el criterio de conformidad está dividido en dos componentes: Metadatos y Lenguaje Común de Intercambio.

- a) **Metadatos:** Los metadatos describen otros datos que explican y facilitan la información importante acerca de estos, lo que permite el acceso a los recursos de información y la reutilización de esta. Una buena gestión de los metadatos debe conducir a una buena calidad de los datos, ya que tenerlos y confiar en ellos puede ayudar a identificar información deficiente, incorrecta y faltante.
  - b) **Lenguaje Común de Intercambio de Información:** El intercambio de información puede realizarse a través de diferentes implementaciones técnicas de servicios de interoperabilidad. La interoperabilidad se define como “capacidad de dos o más sistemas o componentes para intercambiar información y utilizar la información intercambiada.” (ISO/25010).
6. **Exactitud:** Datos diligenciados correctamente. Asegúrese de que los datos estén diligenciados correctamente y que su vez brinden información exacta. Asegúrese de que los datos estén diligenciados correctamente para que cumplan con el criterio de exactitud. A continuación, se describen los tipos de datos soportados actualmente por el portal de datos abiertos de Colombia<sup>1</sup>:
- **Datos numéricos:** Los datos numéricos son aquellos que contienen números, dinero y porcentajes. Se utilizan para representar cantidades o medidas dentro de un conjunto de datos.

Figura No. 2: Datos numéricos



Fuente: Guía de estándares, MINTIC.

- **Fecha y hora:** Los campos de fecha y hora deben de ser diligenciados mediante el formato de la norma ISO 8601<sup>2</sup>. Diligéncielos teniendo en cuenta los lineamientos.

<sup>1</sup> <https://support.socrata.com/hc/en-us/articles/202949918-Importing-Data-Types-and-You->

<sup>2</sup> <https://www.iso.org/iso-8601-date-and-time-format.htm>



Figura No. 3: Formato de fechas básicas

Campo	Formato	Ejemplo
	YYYY-MM-DD	2017-11-05
	YYYY-MM-DDT hh:mm:ss	2017-11-05T 17:37:57
	hh:mm:ss	20:45:50

Fuente: Guía de estándares, MINTIC.

- **Ubicación geográfica:** se define utilizando la notación de grados decimales para latitud y longitud en el siguiente formato: (xx.xxxx, -yy.yyyy) donde “xx.xxxx” es la latitud y “-yy.yyyy” es la longitud.
- **Enlaces URL:** Solo se aceptan tres formatos de URL: FTP, HTTP Y HTTPs. No se aceptan guiones bajos en las URL. Ingrese los enlaces a páginas web en uno de los siguientes formatos de entrada:
  - `<a href="http://www.socrata.com/">Socrata</a>`
  - `http://www.socrata.com/`
  - `Socrata (http://www.socrata.com)`
- **Email:** Todo conjunto de datos debe contener un email de la persona o la entidad encargada de subir la información en la web. Utilice uno de los siguientes formatos de entrada para ingresar los datos de correos electrónicos:
  - `<a href="mailto:sam.gibson@socrata.com">Sam Gibson</a>`
  - `sam.gibson@socrata.com`
  - `Sam Gibson <sam.gibson@socrata.com`
- **Lógico/ binario:** Se utiliza para representar valores binarios y lógicos, tales como 1/0 y Verdadero/Falso. En este campo tenga en cuenta dos (2) posibles valores que serán representados por uno y cero (1 y 0). Tenga en cuenta las siguientes consideraciones a la hora de utilizar los diferentes tipos de datos:
  - No utilice más de un tipo de dato en la misma columna.
  - En los campos numéricos, adopte codificación numérica para los casos especiales, por ejemplo: -88: “No aplica”, -89: “No disponible”, -98: “No sabe”, -99: “No responde”, -77: “Otro”. Esto con el fin de no combinar tipo de datos de texto con datos numéricos.
  - Si el contenido de un campo depende de una etiqueta recurrente, asegúrese de que

esta se utilice de manera uniforme en todo el conjunto de datos. Por ejemplo, si se requiere usar la etiqueta SI/NO en un campo, no mezclar etiquetas como Sí, sí, No, no.

7. **Compleitud:** Campos completamente diligenciados. Para que un conjunto de datos cumpla a cabalidad con los criterios de calidad establecidos en el presente documento de instrucciones, es fundamental que los datos se presenten completos para el uso de los usuarios. Para evitar errores, tenga en cuenta que:
- Que el conjunto de datos tenga campos incompletos y/o vacíos.
  - Que el conjunto de datos tenga registros en blanco.
  - Que se carguen archivos en formatos cerrados que no corresponden a conjuntos de datos tales como formatos .pdf, .doc, .gif, .jpg, .ppt, entre otros.
  - Enlaces a sistemas de información que no permitan la descarga directa de un conjunto de datos en formato abierto: csv, xls,xlsx, json, kml, kmz y zip (shapefile de ESRI).
  - Que el conjunto de datos contenga datos personales publicados sin contar con la autorización de los titulares de acuerdo con la normatividad vigente: Ley 1581 de 2012 – Protección de datos personales.
  - Que el conjunto de datos incluya información clasificada y/o reservada de acuerdo con lo establecido en la Ley 1712 de 2014 - Transparencia y acceso a la información pública.
  - Enlaces rotos de los metadatos y/o conjuntos de datos.
  - Que el conjunto de datos esté compuesto de una única columna.
  - Que el conjunto de datos no cuente con ninguna información en formato de columnas y/o filas.

Tenga en cuenta las siguientes consideraciones a la hora de diligenciar un registro:

- En los diccionarios de datos o etiquetas de campo tipo texto, incluya siempre opciones como “No aplica”, “Otro”, “No sabe”, “No responde”, “No disponible”, según se requiera. Esto con el fin de no dejar espacios vacíos cuando no se tenga la información.
  - No publique conjuntos de datos que contengan una sola columna.
  - No publique conjuntos de datos incompletos.
  - Si existen datos sobre un mismo tema distribuidos en diferentes fuentes de información, consolide un único conjunto de datos que represente la información lo más completa posible para los usuarios.
  - Cuando los conjuntos de datos provienen de un modelo de datos relacional, tenga en cuenta que cada tabla no es un conjunto de datos.
8. **Consistencia:** Datos coherentes y libres de contradicción. Sus datos son consistentes cuando están libres de contradicción y son coherentes respecto a otros datos en el mismo contexto de uso. Esto quiere decir que se utilice la misma codificación de variables y etiquetas en todos los conjuntos de datos. Ejemplo: si se utiliza codificación como 1: Mujer, 2: Hombre, para la variable “Sexo”, utilizar siempre estas convenciones en todos los conjuntos de datos. Aplica también, y con mayor rigurosidad, para las codificaciones que están estandarizadas, como los códigos de la división

político-administrativa del país.

Figura No. 4: Ejemplo de conjunto de datos con errores de consistencia

Id_persona	Sexo	Edad
1	H	12
2	mujer	23
3	Hombre	43
4	hombre	15
5	Mujer	65

Fuente: Guía de estándares, MINTIC

Figura No. 5: Ejemplo de conjunto de datos después de corregir errores

Id_persona	Sexo	Edad
1	Hombre	12
2	Mujer	23
3	Hombre	43
4	Hombre	15
5	Mujer	65

Fuente: Guía de estándares, MINTIC

9. **Precisión:** Nivel de desagregación en que están publicados los datos con respecto al nivel de desagregación en que fueron generados. Los conjuntos de datos se deben publicar con el más alto nivel de desagregación posible, es decir, que se publiquen siempre desde la fuente primaria de la información. Evite hacer agregaciones ya que puede causar que se sesguen algunas dimensiones asociadas.
10. **Portabilidad:** Formatos sin restricciones para la reutilización de los datos. Se trata de la capacidad del conjunto de datos para ser transferido y adaptarse a diferentes formatos, es decir, ser reutilizado. Se refiere además a la facilidad con la que el conjunto de datos se puede procesar, acceder y utilizar fácilmente descargándolo o consumiéndolo a través de interfaces de programación de aplicaciones. Para el caso de los aplicativos, sistemas de información y demás herramientas en donde se almacena la información de la SDG, la DTI cuenta para su consulta con el procedimiento GDI-TIC-P002 Gestión de Sistemas Información.
11. **Credibilidad:** Da cuenta del grado en el cual el conjunto de datos posee información que pueda

ser catalogada como veraz y confiable por los usuarios. Esto se puede estimar con la presencia de información en los metadatos sobre fuentes de información, documentación, normatividad, origen y/o entidad publicadora de los datos.

12. **Comprensibilidad:** Características que permiten al usuario leer e interpretar los datos. Aquí, se evalúan atributos como el nombre del conjunto de datos, descripción, diccionario de datos e índice de etiquetas. En el conjunto de datos se evalúa la claridad en los nombres de los campos y las etiquetas.

Para garantizar el cumplimiento del criterio de comprensibilidad, siga los pasos que se detallan a continuación:

- Asegúrese de que el título del conjunto de datos sea comprensible y completo, respondiendo como mínimo a las siguientes preguntas: ¿qué?, ¿dónde?, ¿cuándo?, ¿vigencia?
- Si utiliza siglas, describa su significado.
- Tenga en cuenta que en la descripción del conjunto de datos deberá ampliar el contexto del título.
- Defina los nombres de los campos de forma clara y comprensible para cualquier usuario.
- Si los nombres de los campos se nombran con un sistema codificado, por ejemplo P1, P2, P3 para referirse a las preguntas de una encuesta, incluya la definición de dichos códigos.

Ejemplo:

- P1: ¿cuántos hermanos tiene?
  - P2: ¿cuál es su ocupación?
  - P3: ¿cuál es su estado civil?
  - Para etiquetas o valores de los campos codificados, incluya la definición de dichas etiquetas
  - Siempre incluya una descripción narrativa de lo que significa cada campo, ejemplo: Población 2018: se refiere a la cantidad de personas estimadas según las proyecciones de población para cada uno de los departamentos de Colombia en el año 2018.
  - Incluya una descripción clara de lo que significa cada registro del conjunto de datos, ejemplo: Cada registro (o fila) de este conjunto de datos representa una localidad de Bogotá y el valor de su proyección poblacional para los próximos 3 años.
13. **Accesibilidad:** Herramientas tecnológicas que garanticen el acceso al usuario que lo requiera, facilitando la disposición, búsqueda y descarga de datos para cualquier persona mediante la disposición de los recursos tecnológicos necesarios para garantizarlo. Los datos deben ser fácilmente visibles y accesibles, poniéndose a disposición sin barreras burocráticas o administrativas que pueden disuadir a las personas de acceder a los datos.
14. **Eficiencia:** Capacidad de la plataforma de realizar análisis y descargas de los datos con unos niveles de desempeño y tiempos esperados. Las plataformas, sistemas de información de la SDG, deben garantizar que los datos puedan ser procesados y proporcionados con los niveles de rendimiento esperados. Además, que cuenten con protocolos que permiten una fácil conexión de los datos dispuestos en la plataforma para ser utilizados en otras aplicaciones directamente y en tiempo real, como pueden ser las herramientas de Business Intelligence, analítica e incluso Excel.

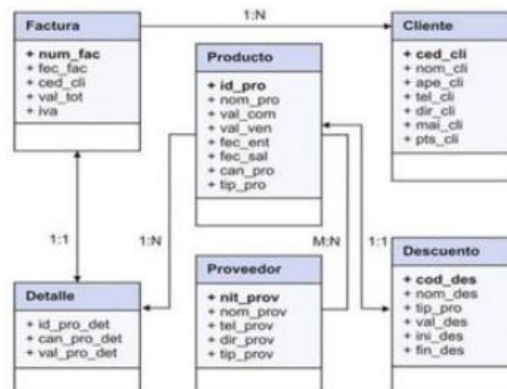
15. **Recuperabilidad:** Atributos que permiten mantener y preservar un nivel específico de operaciones y de calidad. incluso en caso de fallos, manteniendo la recuperabilidad de los datos. es importante que la entidad tenga copias de seguridad de toda la información que maneja y carga en los diferentes aplicativos, plataformas, sistemas de información, etc, y no lo tome como la fuente primaria de almacenamiento.
16. **Disponibilidad:** Garantiza que los usuarios autorizados tengan acceso a la información y a otros activos de información asociados en el lugar, momento y forma en que es requerido. Una plataforma segura debe mantener la información disponible para los usuarios, utilizando estándares de conectividad e interfaces.

Es importante mencionar que este instructivo contiene propuestas y conceptos muy generales que pueden ser aplicables a las diferentes estructuras de datos de las áreas o dependencias de la SDG. Sin embargo, se debe dejar la claridad que existen distintas metodologías de generación de datos lo cual implica que el análisis para evaluar la calidad del dato sea distinto.

Teniendo en cuenta lo anterior, se realizan algunas propuestas que dan un lineamiento y definen criterios generales para mejorar la calidad de los datos, como se muestra a continuación:

1. **Elaborar un manual de estructura de bases de datos organizadas** en tablas que se comunican entre sí por medio de una variable denominada llave (modelo estrella).

Ilustración 1



Una de las ventajas utilizando el modelo estrella, es la administración de un mayor volumen de datos, mayor seguridad de la información y evita la duplicidad de registros.

2. **Sistemas de Información para dar directrices de calidad del dato:** Dentro de la SDG, se pueden utilizar los sistemas de información propios de la entidad, creando formularios con listas desplegables, que permitan minimizar los errores, como por ejemplo, estar por fuera del rango,

duplicidad de registros, error en la digitación, entre otros, con el propósito de dar cumplimiento a los mínimos requeridos de la calidad del datos (Exactitud, oportunidad, accesibilidad, completitud, consistencia, integridad, validez).

3. **Elaborar un flujo de datos:** Esta propuesta hace referencia a generar un flujo de datos, con unas fases, que permiten cumplir con los mínimos establecidos para la calidad del dato, como se ilustra a continuación:

Ilustración 2



Fuente: Elaboración de protocolos de calidad, Ministerio del Medio Ambiente

- **Carga de datos:** Se identifican las principales entradas y salidas del proceso de los datos a través de la visión general de sistema y de lo que se requiere construir.
  - **Validación de datos:** En cuanto a esta fase la entrada de este flujo, se recoge o se obtienen nuevos datos de diferentes fuentes. Para su procesamiento es importante tener en cuenta: 1) perfilar los datos para resaltar los errores 2) ejecutar técnicas de análisis, limpieza y normalización de datos para lograr una visión coherente 3) coincidencia de registros que pertenecen a la misma entidad 4) fusionar los registros para eliminar la información innecesaria y conseguir una única fuente de verdad y finalmente su salida es el almacenamiento de la única fuente de verdad en la fuente destino.
  - **Análisis estadístico de los datos:** Se deben tener técnicas que permitan interpretar los datos numéricos que disponemos, con el objetivo de tomar las decisiones de negocio más eficaces. De hecho, las entidades pueden tomar decisiones 5 veces más rápido si las basan en el análisis de datos (Bain & Company).
  - **Construcción de indicadores:** En cuanto a la construcción de indicadores existen diferentes metodologías. Sin embargo, se recomienda basarse en la guía para la construcción y análisis de indicadores del Departamento Nacional de Planeación-DNP
  - **Publicación:** Se realiza su publicación para su visualización por medio ya sea de tableros de control de acuerdo con lo establecido en el instructivo GCN-IN005-Instrucciones para el diseño de tableros Power BI y en la ficha GCN-F013- Formato fichas de tableros de control o por medio de infografías, boletines, informes o entre otras.
4. **Elaborar un marco para asegurar la calidad de los datos:** En esta propuesta el marco para asegurar la calidad del dato, se refiere a los temas y subtemas que debe cumplir el conjunto de

datos, teniendo en cuenta que estos hacen referencia a los atributos que son exigibles y requeridos para contar con la calidad de los datos:

No.	Tema	Subtema
1.	Prerrequisito de la calidad	<ul style="list-style-type: none"> <li>• Consideraciones de disposiciones legales y normativas</li> <li>• Prerrequisitos</li> <li>• Estructura de los datos</li> <li>• Verificación de cumplimiento de indicadores</li> </ul>
2.	Aseguramiento de la calidad de la información	<ul style="list-style-type: none"> <li>• Catálogo de Componente de Información SDG</li> <li>• Código de Buenas prácticas (DANE)</li> </ul>
3.	Aspectos Metodológicos	<ul style="list-style-type: none"> <li>• Fuentes de información de los datos</li> <li>• Conceptos y definiciones</li> <li>• Metodología de levantamiento de los datos</li> </ul>
4.	Relevancia de los datos	<ul style="list-style-type: none"> <li>• Importancia de los datos</li> <li>• Precisión de los datos: Técnicas estadísticas</li> </ul>
5.	Oportunidad de los datos	<ul style="list-style-type: none"> <li>• Accesibilidad y oportunidad de la información</li> </ul>
6.	Puntualidad	<ul style="list-style-type: none"> <li>• Fechas importantes</li> <li>• Entrega</li> <li>• Publicación</li> </ul>
7.	Accesibilidad	<ul style="list-style-type: none"> <li>• Metadatos: Fichas técnicas y hojas metodológicas</li> </ul>



### 3. DOCUMENTOS RELACIONADOS

#### 3.1 Documentos internos

Código	Documento
GDI-TIC-P002	Gestión de Sistemas de Información
GCN-P006	Herramientas para uso y apropiación y analítica institucional
GCN-P009	Procedimiento gestión de información estadística
GCN-IN005	Instrucciones para el diseño de tableros Power BI
GCN-F013	Formato fichas de tableros de control

#### 3.2 Normatividad vigente

Norma	Año	Epígrafe	Artículo(s)
Decreto Nacional 1499	2017	Por el cual se modifica el Decreto 1083 de 2015, Decreto único reglamentario del sector función pública, en lo relacionado con el Sistema de Gestión, establecido en el artículo 133 de la ley 1753 de 2015	Toda la norma
Ley 1712	2014	Ley de Transparencia y del Derecho de Acceso a la Información Pública Nacional	Toda la norma
Conpes 3920	2018	Política Nacional de Explotación de Datos (Big Data)	Toda la norma
Decreto 454	2020	Por medio del cual se modifica el Decreto 1083 de 2015, Decreto Único Reglamentario del Sector Función Pública, con la incorporación de la política de gestión de la información estadística a las políticas de gestión y desempeño institucional	Toda la Norma
Decreto 1083	2015	Por el cual se expide el Decreto Único Reglamentario del Sector Función Pública	Artículo 2.1.1.1 Artículo 2.1.1.2 Artículo 2.2.22.1.1 Artículo 2.2.22.2.1
	2020	Por la cual se actualizó y publicó la norma técnica para garantizar la calidad del proceso de producción y difusión de las estadísticas oficiales.	Toda la Norma



Norma	Año	Epígrafe	Artículo(s)
Resolución 0783	2018	Por la cual se crea el Comité Institucional de Gestión y Desempeño y se dictan otras disposiciones en la Secretaria Distrital de Gobierno	Toda la Norma
Resolución 0236	2019	Por la cual se modifica la Resolución 0783 de 2018	Toda la Norma

### 3.3. Documentos externos

Nombre	Fecha de publicación o versión	Entidad que lo emite	Medio de consulta
Guía para la implementación de la gestión del conocimiento y la innovación en el marco del modelo integrado de planeación y gestión (MIPG)	Diciembre 2020	Departamento Administrativo de la Función Pública	<a href="https://www.funcionpublica.gov.co/web/e-va/biblioteca-virtual/-/document_library/bGsp2IjUBdeu/view_file/37827592">https://www.funcionpublica.gov.co/web/e-va/biblioteca-virtual/-/document_library/bGsp2IjUBdeu/view_file/37827592</a>
Manual Operativo del Modelo Integrado de Planeación y Gestión MIPG, Consejo para la Gestión y Desempeño Institucional Versión 4	Marzo de 2021	Departamento Administrativo de la Función Pública	<a href="https://www.funcionpublica.gov.co/documents/28587410/34112007/Manual+Operativo+MIPG.pdf/ce5461b4-97b7-be3b-b243-781bbd1575f3">https://www.funcionpublica.gov.co/documents/28587410/34112007/Manual+Operativo+MIPG.pdf/ce5461b4-97b7-be3b-b243-781bbd1575f3</a>
Guía G. INF. 06 Guía Técnica de Información	2014	Ministerio de Tecnologías de la Información y las Comunicaciones MinTic	<a href="https://www.mintic.gov.co/arquitecturati/630/articles-9258_recurso_pdf.pdf">https://www.mintic.gov.co/arquitecturati/630/articles-9258_recurso_pdf.pdf</a>
NTC PE 1000	2020	ICONTEC	Digital