República de Colombia

Gobernación de Santander

POLÍTICA DE CALIDAD DE DATOS Y PLAN DE CALIDAD DE DATOS

TABLA DE CONTENIDO

NORMATIVIDAD APLICABLE 5
DEFINICIONES 6
1. INTRODUCCIÓN 9
1.1. Objetivo 10
1.2. Alcance 10
2. DECLARACIÓN DE LA POLÍTICA 11
2.1. Principios de Calidad de Datos 11
2.2. Disponibilidad 12
2.3. Usabilidad 12
2.3.1 Definición/Documentación: Los datos deben estar bien documentados, con definiciones claras para que los usuarios sepan cómo interpretarlos. 13
2.4. Confiabilidad 13
2.5. Relevancia 13
2.6. Calidad de presentación 13
3. ROLES Y RESPONSABILIDADES 14
4. MEJORA DE LA CALIDAD 16
4.1. Técnicas para el mejoramiento de la calidad del Dato 16
4.2. Procesos para el mejoramiento de la calidad del Dato 17
5. ERRORES DE CALIDAD 19
6. Cumplimiento 24
6.1. Indicador de Actualidad 24
6.2. Indicador de Trazabilidad 25
6.3. Indicador de Completitud 25
6.4. Indicador de Consistencia 26
6.5. Indicador de Credibilidad 27
7. SELLO DE EXCELENCIA 29
8. PLAN DE CALIDAD DE DATOS 30
8.1. PLAN DE TRABAJO CALIDAD DE DATOS 31
9. BIBLIOGRAFÍA 33

TABLA DE ILUSTRACIONES

Ilustración 1 Calidad de Producto de Datos 11
Ilustración 2 Estandar universal de calidad de datos de dos capas 12
Ilustración 3 Técnicas mejoramiento de la calidad del Dato 16
Ilustración 4 Procesos para el mejoramiento de la calidad del Dato 17
Ilustración 5 Sello de Excelencia Gobierno Digital 29

TABLA DE TABLAS

Tabla 1 Errores de calidad y de Publicación 19
Tabla 2 Reglas para el cálculo de indicadores de actualidad. 24
Tabla 3 Reglas para el cálculo de indicadores de trazabilidad 25
Tabla 4 Reglas para el cálculo de indicadores de trazabilidad 26
Tabla 5 Ilustración 9 Reglas para el cálculo de indicadores de consistencia 26
Tabla 6 Reglas para el cálculo de indicadores de comprensibilidad 27
Tabla 7 PLAN DE TRABAJO 31

NORMATIVIDAD APLICABLE

Conpes 3920 de 2018 Política Nacional de Explotación de Datos (Big Data).

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.

Decreto 103 de 2015 Por el cual se reglamenta parcialmente la Ley 1712 de 2014 y se dictan otras disposiciones.

Decreto 1078 de 2015 Por medio del cual se expide el Decreto Único Reglamentario del Sector de Tecnologías de la Información y las Comunicaciones.

Decreto 415 de 2016 Por el cual se adiciona el Decreto Único reglamentario del sector de la Función Pública, Decreto 1083 de 2015, en lo relacionado con la definición de los lineamientos para el fortalecimiento institucional en materia de Tecnologías de la Información y las Comunicaciones.

Ley 1581 de 2012 Por el cual se dictan disposiciones generales para la Protección de Datos Personales.

Ley 1712 de 2014 Por medio de la cual se crea la Ley de Transparencia y del Derecho de Acceso a la Información Pública Nacional y se dictan otras disposiciones.

Ley 1757 de 2015 Por la cual se dictan disposiciones en materia de Promoción y Protección del Derecho a la Participación Democrática.

Resolución 1519 de 2020 Definen los estándares y directrices para publicar la información señalada en la Ley 1712 del 2014) y se definen los requisitos materia de acceso a la información pública, accesibilidad web, seguridad digital, y datos abiertos.

Resolución 3564 de 2015 Por el cual se reglamenta los estándares de publicación y divulgación de la información

Resolución 460 de 2022 Plan Nacional de Infraestructura de Datos (PNID) y su Hoja de Ruta.

DEFINICIONES

Arquitectura de datos e información: La arquitectura de datos e información es la estructura organizativa que define cómo se recolectan, almacenan, gestionan y distribuyen los datos en una organización. En la Gobernación, esta arquitectura debe garantizar que los datos sean accesibles y utilizables de manera eficiente, permitiendo la integración y el intercambio de información entre diferentes departamentos y sistemas.

Calidad de datos: La calidad de los datos se refiere a qué tan adecuados y confiables son los datos para su uso en procesos de toma de decisiones. Los datos de calidad son precisos, completos, consistentes, actualizados y accesibles. En la Gobernación, esto es esencial para tomar decisiones correctas y eficientes en temas que impactan a la población, como la asignación de recursos y la prestación de servicios.

Dato: Un dato es un valor específico que representa una característica o atributo de una entidad. Puede ser un número, texto o cualquier otro tipo de información. Por ejemplo, en la Gobernación, un dato podría ser el nombre de un ciudadano, su dirección o su número de identificación.

Datos abiertos: Los datos abiertos son aquellos datos que se ponen a disposición del público de manera libre, sin restricciones de acceso, con el fin de fomentar la transparencia y permitir el uso y reutilización por parte de ciudadanos, investigadores y empresas. En la Gobernación, los datos abiertos pueden incluir información sobre presupuestos, proyectos y estadísticas que cualquier persona pueda consultar y analizar.

Dato maestro: Un dato maestro es un tipo de dato crítico y central para el funcionamiento de una organización, que se utiliza en múltiples áreas y sistemas. Los datos maestros suelen incluir información fundamental sobre ciudadanos, proveedores o ubicaciones geográficas. En la Gobernación, por ejemplo, los datos maestros pueden incluir el registro único de ciudadanos o las ubicaciones oficiales de oficinas gubernamentales.

Gobierno de datos: El gobierno de datos es el conjunto de políticas, procesos y estándares que aseguran la gestión adecuada de los datos dentro de una organización. Su objetivo es garantizar que los datos sean confiables, seguros y cumplan con los estándares de calidad establecidos. En la Gobernación, el gobierno de datos incluye la gestión de cómo se recopilan, almacenan y utilizan los datos en todos los departamentos.

Metadato: Un metadato es un dato que describe otros datos. Proporciona información sobre la estructura, contenido, calidad y contexto de un conjunto de datos, facilitando su organización y búsqueda. En la Gobernación, los metadatos ayudan a catalogar la información en las bases de datos para que pueda ser fácilmente accesible y comprensible por los usuarios.

Gobernanza: La gobernanza es el marco general que define cómo se toman decisiones estratégicas y operativas en una organización, asegurando el cumplimiento de políticas y regulaciones. En el contexto de la calidad de los datos, la gobernanza garantiza que existan

reglas claras para la administración de los datos, alineadas con los objetivos institucionales de la Gobernación.

Información: La información es el conjunto de datos procesados y organizados de manera que tengan un significado y puedan ser utilizados para la toma de decisiones. En la Gobernación, la información se utiliza para evaluar situaciones, planificar políticas públicas y administrar recursos, por lo que su calidad es fundamental.

Interoperabilidad: La interoperabilidad es la capacidad de diferentes sistemas y organizaciones para trabajar juntos, permitiendo el intercambio y uso de información entre ellos. En la Gobernación, la interoperabilidad es esencial para que los departamentos y entidades puedan compartir datos de manera eficiente, asegurando la coordinación en la gestión pública.

Seguridad de la información: La seguridad de la información es el conjunto de prácticas y tecnologías utilizadas para proteger los datos contra accesos no autorizados, alteraciones o destrucción. En la Gobernación, es crucial para proteger la información personal de los ciudadanos y asegurar que los datos críticos estén resguardados contra amenazas como el robo, la manipulación o la pérdida.

CONTROL DE CAMBIOS				
VERSIÓN	**FECHA**	**DESCRIPCIÓN DEL CAMBIO**	**REVISÓ**	**APROBÓ**
0	24/11/2023	Creación Del Documento	Cristhian Becerra Hernandez	Miguel Guillermo Sarmiento
1	09/01/2025	Se ajusta la política a los criterios de calidad que menciona la Guía de estándares de calidad e interoperabilidad de los datos abiertos del gobierno de Colombia	Sandra Liliana Bautista Escobar	Shirley Paola Castellanos Martínez

1. INTRODUCCIÓN

Este documento ha sido desarrollado con el propósito fundamental de establecer los lineamientos necesarios para garantizar la calidad de los datos y la información gestionada dentro de la Gobernación de Santander. Si bien no pretende ser un manual exhaustivo que abarque todas las posibles situaciones relacionadas con la gestión de datos, constituye una base sólida y un punto de partida esencial para la identificación e implementación de buenas prácticas en el manejo de la información. La finalidad última de estos lineamientos es asegurar que la entidad cuente con datos de alta calidad que permitan generar valor público tangible para los ciudadanos, facilitando la prestación de servicios más eficientes y la toma de decisiones mejor fundamentada. Este documento refleja el compromiso de la Gobernación con la excelencia en la gestión de la información como recurso estratégico para el desarrollo regional.

1.1. Objetivo

Establecer lineamientos normativos que permitan realizar la gestión de datos producidos en la Gobernación de Santander con la finalidad de garantizar la calidad de estos y asegurar su disponibilidad y valor al ciudadano, mejorando los procesos internos y creando una cultura organizacional centrada en el dato que permita la creación de componentes de información que serán la fuente para crear servicios de información fundamentados en tecnología y habilitará a la administración tomar decisiones asertivas respaldados en la analítica de los datos

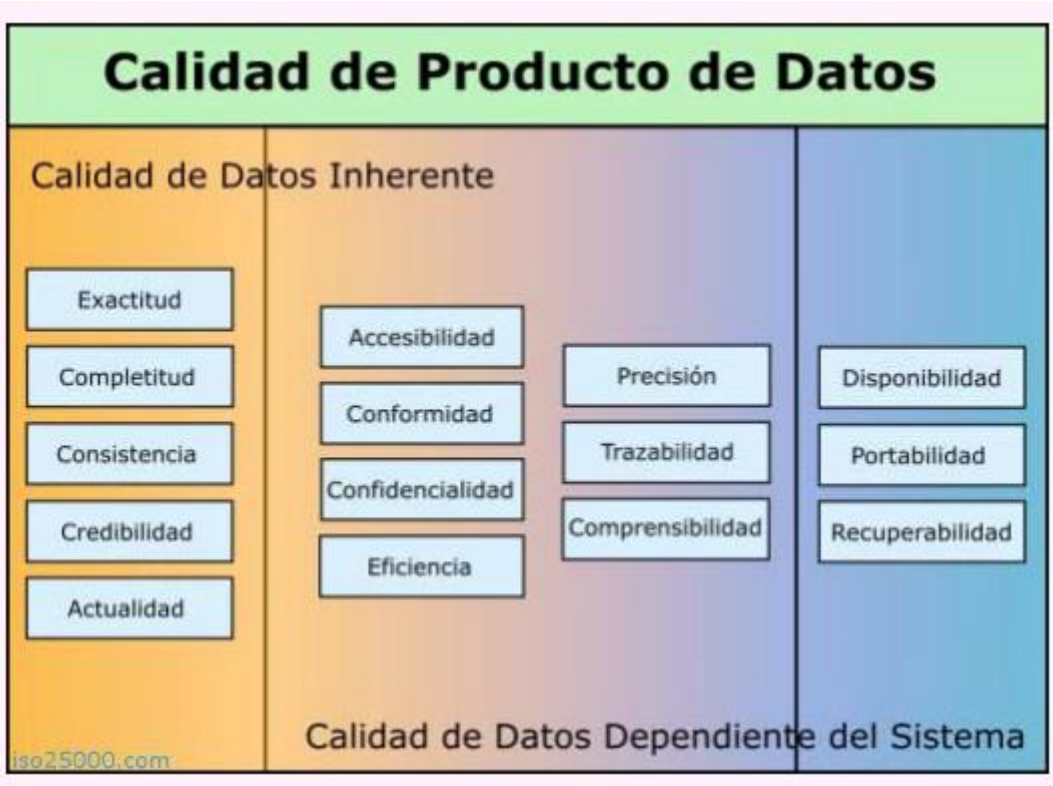
1.2. Alcance

El alcance de esta política ha sido diseñado para abarcar de manera comprehensiva todos los aspectos críticos relacionados con la gestión de datos en la Gobernación. En primer lugar, la política aplica a todos los empleados de la entidad que, en el desempeño de sus funciones, tengan cualquier nivel de interacción con datos institucionales, desde su recolección hasta su procesamiento y análisis. Asimismo, se extiende a todos los sistemas y aplicaciones que procesan datos dentro de la organización, incluyendo tanto las plataformas existentes como aquellas que se implementen en el futuro.

2. DECLARACIÓN DE LA POLÍTICA

2.1. Principios de Calidad de Datos

Ilustración 1 Calidad de Producto de Datos



Tomado de DATOS ABIERTOS Y CALIDAD DEL DATO - MINTIC

La Calidad del Producto de Datos se puede entender como el grado en que los datos satisfacen los requisitos definidos por la organización a la que pertenece el producto. Son precisamente estos requisitos los que se encuentran reflejados en el modelo de Calidad de Datos mediante sus características (Exactitud, Completitud, Consistencia, Credibilidad, Actualidad, Accesibilidad,). (Fuente: ISO/IEC 25012)

Ilustración 2 Estándar universal de calidad de datos de dos capas

Tomado de DATOS ABIERTOS Y CALIDAD DEL DATO - MINTIC

2.2. Disponibilidad

Se refiere a la facilidad con la que los datos pueden ser accedidos y utilizados cuando se necesitan. Abarca:

2.2.1 **Accesibilidad**: Capacidad para obtener los datos cuando son necesarios, sin barreras que impidan su uso.

2.2.2 **Oportunidad**: Los datos están disponibles en el momento adecuado, es decir, cuando se requieren para la toma de decisiones.

2.2.3 **Autorización**: Asegura que solo las personas con permisos adecuados pueden acceder a los datos, protegiendo la información sensible.

2.3. Usabilidad

Se refiere a qué tan fácil es para los usuarios entender y trabajar con los datos. Involucra:

2.3.1 Definición/Documentación: Los datos deben estar bien documentados, con definiciones claras para que los usuarios sepan cómo interpretarlos.

2.3.4 Credibilidad: Los datos deben ser confiables y considerados correctos por los usuarios que los manejan.

2.3.4 Metadatos: Información que describe otros datos, permitiendo que los usuarios comprendan el contexto, la estructura y el significado de los datos.

2.4. Confiabilidad

Hace referencia a la confianza que se puede tener en que los datos son precisos y consistentes. Incluye:

2.4.1 Precisión: Los datos deben reflejar correctamente la realidad que representan.

2.4.2 Integridad: Los datos deben estar completos, sin elementos faltantes ni errores.

2.4.3 Consistencia: Los datos deben ser uniformes y coherentes a través de diferentes bases de datos o sistemas.

2.4.4 Completitud: Garantiza que los datos no tengan omisiones ni estén incompletos.

2.4.5 Auditabilidad: Los datos deben poder rastrearse, verificarse y revisarse a través de auditorías.

2.5. Relevancia

Se refiere a la capacidad de los datos para ser útiles y adecuados para el propósito para el cual se recopilan o utilizan:

2.5.1 Aptitud: Los datos deben ser apropiados y adecuados para cumplir con las necesidades de la organización o el análisis que se va a realizar.

2.6. Calidad de presentación

La forma en que los datos se presentan influye en su facilidad de uso y comprensión. Abarca:

2.6.1 Legibilidad: Los datos deben ser fáciles de leer y comprender.

2.6.2 Estructura: Los datos deben tener un formato organizado y consistente que permita su uso y análisis de manera eficiente.

3. ROLES Y RESPONSABILIDADES

Propietarios de Datos: Son las personas o grupos responsables de la gestión de conjuntos específicos de datos. Esto incluye la responsabilidad sobre la calidad, integridad y uso adecuado de esos datos.

Responsabilidades:

Aplicar políticas y estándares adoptados por el departamento para la gestión de los datos bajo su propiedad.

Asegurarse de que los datos sean precisos, completos y actualizados.

Aprobar accesos y permisos a los datos, asegurando que solo los usuarios autorizados puedan acceder a la información.

Colaborar con otros departamentos para garantizar que los datos sean utilizados de manera eficiente y efectiva.

Custodios de Datos: Son responsables del mantenimiento diario y la gestión operativa de los conjuntos de datos. Se aseguran de que los datos se manejen de acuerdo con las políticas establecidas por los propietarios de datos.

Responsabilidades:

Monitorear la calidad de los datos y realizar auditorías regulares para identificar y corregir errores o inconsistencias.

Implementar procesos de validación y limpieza de datos para mantener la integridad de la información.

Asegurarse de que los sistemas de almacenamiento de datos sean seguros y estén disponibles para los usuarios autorizados.

Proporcionar soporte técnico y capacitación a los usuarios sobre el manejo adecuado de los datos.

Usuarios de Datos: Son los individuos o equipos que utilizan los datos para llevar a cabo sus tareas y responsabilidades en la Gobernación. Esto incluye la entrada, análisis y reporte de datos.

Responsabilidades:

Seguir los procedimientos y estándares establecidos para el manejo de datos, garantizando su correcta utilización.

Informar a los propietarios y custodios de datos sobre cualquier inconsistencia, error o problema encontrado en los datos.

Participar en capacitaciones y talleres sobre mejores prácticas en la gestión de datos.

Utilizar los datos de manera ética y responsable, respetando la privacidad y confidencialidad de la información.

4. MEJORA DE LA CALIDAD

4.1. Técnicas para el mejoramiento de la calidad del Dato

Ilustración 3 Técnicas mejoramiento de la calidad del Dato

Tomado de DATOS ABIERTOS Y CALIDAD DEL DATO - MINTIC

Limpieza de Datos: Proceso de identificar y corregir errores o inconsistencias en los datos. Esto incluye eliminar duplicados, corregir errores tipográficos y asegurar que los datos cumplan con los formatos establecidos.

Validación de Datos: Proceso de verificar que los datos ingresados cumplan con ciertos criterios o reglas. Esto ayuda a asegurar que la información sea precisa y confiable desde el momento de su entrada.

Manipulación de Datos: Involucra la transformación, combinación o restructuración de los datos para facilitar su análisis y presentación. Esto incluye operaciones como agregar, dividir o resumir datos.

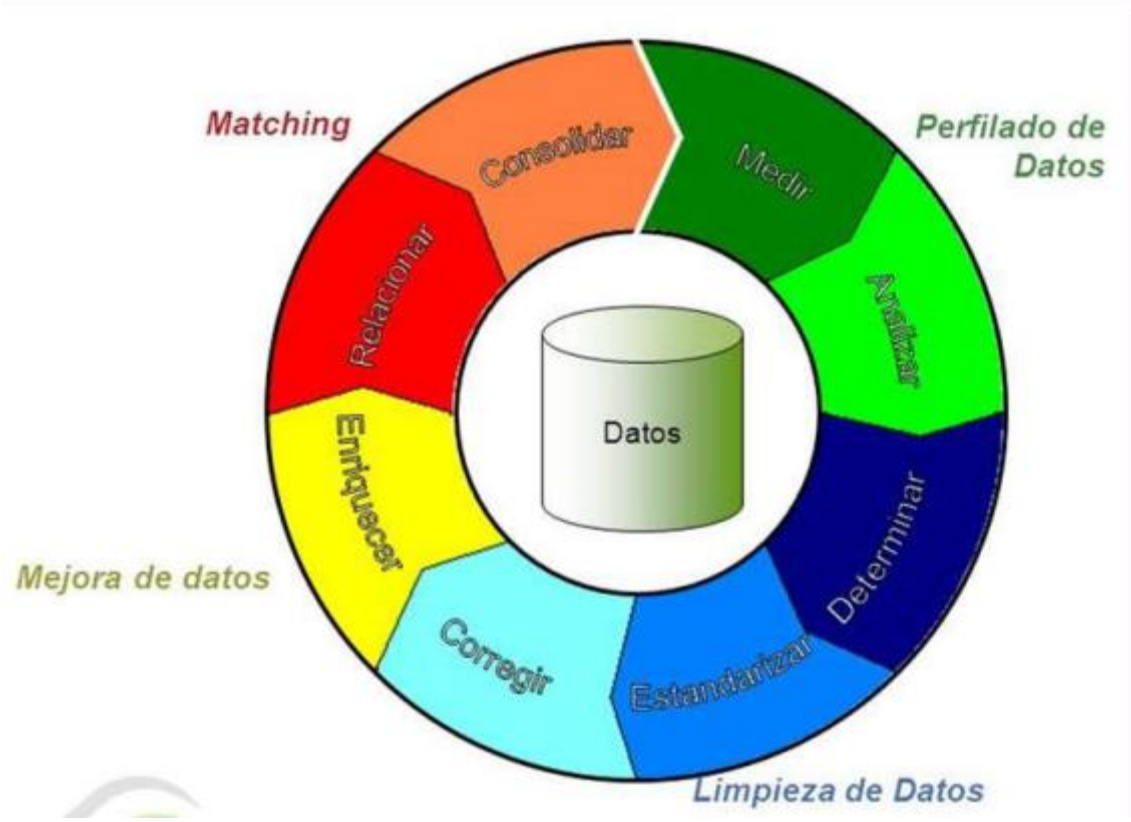
Pruebas de Calidad de Datos: Proceso de realizar pruebas para evaluar diferentes aspectos de la calidad de los datos, como precisión, completitud, y consistencia. Se utilizan métricas específicas para identificar áreas de mejora.

Perfilado de Datos: Análisis de los datos existentes para entender su estructura, calidad y contenido. Implica la recolección de estadísticas y métricas sobre los datos para identificar patrones, valores atípicos o problemas de calidad.

Filtrado de Datos: Proceso de seleccionar y extraer subconjuntos específicos de datos de un conjunto más grande, basado en criterios determinados. Esto ayuda a enfocar el análisis en información relevante y útil.

4.2. Procesos para el mejoramiento de la calidad del Dato

Ilustración 4 Procesos para el mejoramiento de la calidad del Dato



Tomado de DATOS ABIERTOS Y CALIDAD DEL DATO - MINTIC

Relacionar: Establecer conexiones entre diferentes conjuntos de datos para obtener una visión más completa y coherente de la información. Esto ayuda a entender cómo los datos se relacionan y complementan entre sí.

Consolidar: Agrupar y combinar datos de diferentes fuentes o sistemas en un único conjunto de datos, lo que facilita su análisis y gestión. La consolidación ayuda a reducir redundancias y mejorar la coherencia.

Medir: Evaluar y cuantificar diferentes aspectos de la calidad de los datos utilizando métricas específicas. Esto permite identificar áreas de mejora y establecer benchmarks de calidad.

Analizar: Examinar los datos para identificar patrones, tendencias o anomalías que puedan afectar su calidad. El análisis permite extraer información útil y tomar decisiones informadas basadas en datos.

Determinar: Establecer criterios y normas para evaluar la calidad de los datos, así como identificar las necesidades de los usuarios y las expectativas en relación con la información.

Estandarizar: Crear y aplicar estándares para la recopilación, almacenamiento y manejo de datos. La estandarización garantiza que todos los datos se manejen de manera consistente, lo que facilita su uso y comparación.

Corregir: Identificar y solucionar errores o inconsistencias en los datos. Este proceso es fundamental para mantener la calidad y la integridad de la información.

Enriquecer: Mejorar los conjuntos de datos existentes añadiendo información adicional que pueda proporcionar más contexto o valor. Esto puede incluir la integración de datos de fuentes externas.

5. ERRORES DE CALIDAD

La clasificación y priorización de errores en la calidad de datos es un proceso crítico dentro del marco de gestión de información pública. Los errores de calidad abarcan problemas fundamentales como la inconsistencia en los valores de los datos, duplicidad de registros, campos incompletos o nulos, los cuales afectan directamente la integridad y confiabilidad de los conjuntos de datos.

Tabla 1 Errores de calidad y de Publicación

Tipo de Error	Descripción Error	Solución
Título y Descripción mal nombrados	El título y descripción del conjunto de datos está incompleto, presenta siglas, caracteres especiales o información que no es clara para los usuarios del conjunto de datos	Se recomienda actualizar el título y la descripción siguiendo las recomendaciones de calidad de las herramientas de apoyo
La Metadata del conjunto de datos está vacía o incompleta	La metadata está incompleta o vacía por lo cual se requiere completar los siguientes campos: {Incluir campos}	Se recomienda actualizar o incluir la información relacionada para los siguientes campos {Incluir campos} en la metadata de su conjunto de datos
Conjunto de datos no es una base de datos o presenta poca información para reutilización	El conjunto de datos tiene muy pocos registros (Menos de 50) lo que no permitiría la reutilización por parte de los ciudadanos para realizar un producto o servicio	Se recomienda AUMENTAR el número de registros del conjunto de datos con el fin de que sea de utilidad para los usuarios del conjunto de datos, agregar nuevas columnas y/o consolidar con otras bases de datos complementarias o históricas.

Los Datos no hacen parte de los conjuntos de datos como mínimos a publicar en el portal de datos abiertos como Instrumentos de gestión de información pública.	Los Datos no hacen parte de los conjuntos de datos como mínimos a publicar en el portal de datos abiertos como Instrumentos de gestión de información pública de acuerdo a lo que establece la ley 1712 de 2014 de transparencia y del derecho de acceso a la información pública	Se recomienda de acuerdo a lo establecido por la ley 1712 publicar como mínimo: REGISTRO DE ACTIVOS DE INFORMACIÓN, ÍNDICE DE INFORMACIÓN CLASIFICADA
El conjunto de datos hace parte de la información a publicar en el portal propio de la entidad en la sección de transparencia	De acuerdo a la Ley 1712 de 2014 y a la Resolución 1519 de 2020, El conjunto de Datos que se está publicando hace parte de la información mínima a publicar en el sitio web propio de la entidad en las sesiones: 1. Información de la entidad 2. Normativa 3. Contratación 4. Planeación, Presupuesto e Informes 5. Trámites 6. Participa 8. Información específica para Grupos de Interés	Se recomienda publicar en el sitio propio de acuerdo a la normativa de acuerdo a los numerales: 1. Información de la entidad 2. Normativa 3. Contratación 4. Planeación, Presupuesto e Informes 5. Trámites 6. Participa 8. Información específica para Grupos de Interés y publicar en datos abiertos lo que dicta el numeral 7. Datos abiertos
Conjunto de datos con campos vacíos y/o basura	El conjunto de datos presenta campos vacíos en las {nombre de las columnas}	Se recomienda hacer uso de herramientas de calidad de datos para estandarizar los campos e identificar los errores presentados.

Conjunto de datos presenta muy pocas columnas	El conjunto de datos tiene muy pocas columnas (Menos de 3) lo que no permitiría la reutilización por parte de los ciudadanos para realizar un producto o servicio	Se recomienda mejorar la completitud del conjunto de datos agregando nuevas columnas con el fin de que sea de utilidad para los usuarios del conjunto de datos
Conjunto de datos presenta columnas mal nombradas o sin nombre (Unnamed Column)	El conjunto de datos presenta un error al nombrar una o varias columnas y al ser publicado quedaron sin nombre (UNNAMED COLUMN)	Se recomienda hacer uso de herramientas de calidad de datos para estandarizar los campos e identificar los errores presentados.
Dato abierto es un subconjunto de un conjunto de dato de una entidad nacional	Este conjunto de datos es un subconjunto de un dato abierto de una entidad nacional que consolida esta informacion.	Se recomienda para evitar datos duplicados y desactualizados en el Portal Nacional de Datos Abiertos, ELIMINAR su conjunto de datos , y CREAR una vista filtrada del conjunto de datos maestro del de la entidad productora.
El conjunto de datos esta desactualizado	El conjunto de datos esta desactualizado y no cumple con los criterios de periodicidad de actualización establecidos en el conjunto de datos en el campo "Frecuencia de Actualización"	Se recomienda ACTUALIZAR el conjunto de dato y mejorar la completitud con el fin de que sea de utilidad para los usuarios del conjunto de datos

Columnas o campos con información nula o vacía	Grado en 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.	Se recomienda completar los campos vacíos o nulos de las columnas que presentan esta inconsistencia, de no ser un campo que no de valor o aporte al contexto del conjunto de datos se recomienda eliminarla.
Conjunto de datos no permite ser interpretado.	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.	Se recomienda mejorar los metadatos incluyendo información o campos para poder clasificar los registros, los datos deben encontrarse en formatos que permitan el procesamiento automático, con el más alto nivel de detalle posible, no en forma agregada ni modificada, así mismo se requiere también un contexto para darles sentido, propósito y uso.
El conjunto de datos contiene datos personales	El conjunto de datos incumple con el atributo de calidad de la confidencialidad	

La información se está duplicando y ya está siendo publicada en el portal en el portal de datos abiertos por otra entidad	La información se está duplicando y ya está siendo publicada en el portal en el portal de datos abiertos por otra entidad que puede pertenecer a el mismo, departamento , municipio , ciudad o región, o área o dependencia .	Se recomienda validar el conjunto de datos nacionales estratégicos y/o contactar la entidad que suministra la misma información ,consoliden y establezcan el rol de la pertinencia de publicación.
Los registros del conjunto de datos se encuentran duplicados.	La unicidad se define como la medida en que todos los valores distintos de un elemento de datos aparecen sólo una vez.	Se recomienda actualizar el conjunto de datos de manera se eliminen registros duplicados, haciendo uso de las herramientas de calidad de datos como open refine
La información contenida en el conjunto de datos, puede no estar considerada como Dato abierto.	El conjunto de datos puede ser no reutilizado por parte de los ciudadanos para realizar un producto o servicio, en cuanto al contexto particular.	Se recomienda ACTUALIZAR el conjunto de dato y mejorar la pertinencia con el fin de que sea de utilidad para los usuarios del conjunto de datos. En caso de que la información solo sea de interés particular de la entidad se recomienda ELIMINAR el conjunto de datos y disponer esta información en la página web de la misma

Adaptado de *Errores Frecuentes de Calidad Conjuntos de Datos Abiertos del Portal*, por datos.gov.co, s.f., https://www.datos.gov.co.

6. Cumplimiento

En el marco del cumplimiento de la calidad de datos, el monitoreo juega un papel fundamental como herramienta de seguimiento continuo. El establecimiento de KPIs (Indicadores Clave de Desempeño) de calidad de datos permite medir y evaluar de manera objetiva aspectos cruciales como la completitud de los registros, la exactitud de la información, la consistencia entre diferentes fuentes de datos, la actualización oportuna y la conformidad con los estándares establecidos por MinTIC.

6.1. Indicador de Actualidad

Reglas de Cálculo: para realizar el cálculo de los indicadores se definieron las siguientes reglas:

Tabla 2 Reglas para el cálculo de indicadores de actualidad.

Criterio	Definición Guía Interoperabilidad	Medible	Métrica por Conjunto de Datos
Actualidad	Vigencia y actualidad de los datos publicados.	Si	Cumple/ No cumple La última fecha de actualización de datos y metadatos se encuentra en el intervalo entre el día de medición, menos el período de frecuencia de actualización definida. Ejemplo: Hoy menos un año para frecuencia anual.

Tomado de Modelo de uso: Calidad de Datos. – MINTIC

Cálculo de indicador:

Paso 1: identificar la frecuencia de la actualización del conjunto de datos, a partir de la información en la metadata.

Paso 2: identificar cuál fue la última fecha de actualización de los datos.

Paso 3: obteniendo la fecha esperada de actualización. Para esto se toma la frecuencia de actualización y se convierte en un valor de día, haciendo la siguiente conversión:

- Frecuencia diaria: valor: 1
- Frecuencia semanal: valor: 7
- Frecuencia quincenal: valor: 15
- Frecuencia mensual: valor: 30
- Frecuencia trimestral: valor: 90
- Frecuencia semestral: valor:183
- Frecuencia anual: valor: 365.

Se toma la fecha actual y se le resta el número de días obtenido en la variable "Valor", dando la fecha mínima de actualización.

Paso 4 (Salida): se compara si la fecha de actualización de los datos es mayor o igual a la fecha esperada de actualización. En caso de ser verdadera la comparación, el indicador será igual a 10, en caso de que sea falso, el indicador será igual a 0.

6.2. Indicador de Trazabilidad

Reglas de Cálculo: para realizar el cálculo de los indicadores se definieron las siguientes reglas:

Tabla 3 Reglas para el cálculo de indicadores de trazabilidad

Criterio	Definición Guía Interoperabilidad	Medible	Métrica por Conjunto de Datos
Trazabilidad	Histórico del conjunto de datos disponible: fechas de creación, publicación y actualizaciones.	Si	Cumple/No cumple Cumple si tiene fecha de actualización de datos, fecha de actualización de metadatos y fecha de creación.

Tomado de Modelo de uso: Calidad de Datos. – MINTIC

Cálculo de indicador

Paso 1: se obtiene la fecha de actualización de datos.

Paso 2: se obtiene la fecha de actualización de Metadatos.

Paso 3: se obtiene la fecha de creación de Metadatos.

Paso 4: si el conjunto de datos no tiene fecha de actualización, el indicador es 0.

Paso 5: si el conjunto de datos no tiene fecha de creación, el indicador es 0.

Paso 6: si la metada no tiene fecha de actualización, el indicador es 0.

Paso 7 (Salida):si el conjunto de datos tiene fecha de creación, fecha de actualización de datos y de metadatos, el indicador es 10.

6.3. Indicador de Completitud

Reglas de Cálculo: para realizar el cálculo de los indicadores se definieron las siguientes reglas:

Tabla 4 Reglas para el cálculo de indicadores de trazabilidad

Criterio	Definición Guía Interoperabilidad	Medible	Métrica por Conjunto de Datos
Completitud	Campos completamente diligenciados.	Si	Porcentaje de datos diligenciados correctamente en el conjunto de datos. Se promedia el resultado de la evaluación hecha a cada columna del conjunto de datos: 1- (cantidad de datos faltantes/ cantidad total de registros)

Tomado de Modelo de uso: Calidad de Datos. – MINTIC

Cálculo de indicador

Paso 1: se calcula el número de registros por cada columna del conjunto de datos.

Paso 2: se suman la cantidad de registros existentes por cada columna.

Paso 3: se calcula el número de registros faltantes por cada columna del conjunto de datos.

Paso 4: se suman la cantidad de registros faltantes por cada columna de datos.

Paso 5 (Salida): se realiza la siguiente operación:

(1- (suma registros faltantes / suma registros totales)) x 10

6.4. Indicador de Consistencia

Reglas de Cálculo: para realizar el cálculo de los indicadores se definieron las siguientes reglas.

Tabla 5 Ilustración 9 Reglas para el cálculo de indicadores de consistencia

Criterio	Definición Guía Interoperabilidad	Medible	Métrica por Conjunto de Datos
Consistencia	Datos coherentes y libres de contradicción.	Si	**1)** Cantidad de columnas a las que les aplica la medición de consistencia, de acuerdo con la existencia de valores de dominio estándar del campo definido. **2)** Porcentaje de columnas consistentes en el conjunto de datos. Se evalúa la consistencia de cada columna: cantidad de columnas consistentes/total de columnas a las que les aplica consistencia

Tomado de Modelo de uso: Calidad de Datos. – MINTIC

Cálculo de indicador

Paso 1: calcular el número de columnas existentes y el número de registros en esas columnas.

Paso 2: calcular el número de columnas repetidas y el número de registros en esas columnas.

Paso 3: calcular el porcentaje de registros repetidos en las columnas de acuerdo a la siguiente formula:

Columnas = *(1 – (total registros repetidos / total registros)) x 10*

Paso 4: calcular el número de filas repetidas y el número de registros en esas filas.

Paso 5: calcular el número de filas repetidas y el número de registros en esas filas.

Paso 6: calcular el porcentaje de registros repetidos en las filas, de acuerdo a la siguiente formula:

Filas = *(1 – (total registros repetidos / numero registros totales)) x 10*

Paso 7 (Salida): cálculo del indicador de consistencia, a través de la siguiente formula:

Indicador Consistencia = *(Columnas + Filas) / 2*

6.5. Indicador de Credibilidad

Reglas de Cálculo: para realizar el cálculo de los indicadores se definieron las siguientes reglas:

Tabla 6 Reglas para el cálculo de indicadores de comprensibilidad

Criterio	Definición Guía Interoperabilidad	Medible	Métrica por Conjunto de Datos
Comprensibilidad	Características que permiten al usuario leer e interpretar los datos.	Parcial	Porcentaje de descripciones de campo del conjunto de datos en la metadata. Cantidad de descripciones/Cantidad de campos totales del conjunto de datos

Tomado de Modelo de uso: Calidad de Datos. – MINTIC

Cálculo de indicador

Paso 1: obtener el número de columnas en la metadata del conjunto de datos y sus nombres.

Paso 2: identificar el número de columnas en el conjunto de datos y sus nombres.

Paso 3: identificar el número de columnas que coinciden de acuerdo a su nombre, haciendo la comparación entre los nombres de las columnas de la metada y los conjuntos de datos.

Paso 4: identificar el número de columnas únicas entre la data y la metadata, de acuerdo a su nombre.

Paso 5: cálcular el indicador a través de la siguiente fórmula:

Comprensibilidad = *(1 – (nombre coincidentes/columnas únicas)) x 10*

7. SELLO DE EXCELENCIA

Los requisitos de calidad de datos dentro del Sello de Excelencia de Gobierno Digital de Colombia establecen estándares específicos que garantizan la excelencia en la publicación y gestión de datos abiertos gubernamentales. Los datos deben cumplir con criterios fundamentales como la completitud, permitiendo que la información esté íntegra y sin campos vacíos relevantes; la actualización oportuna, asegurando que los datos reflejen la realidad temporal que representan; la precisión y exactitud de la información publicada; y la consistencia en formatos y estructuras. Además, se exige que los conjuntos de datos estén documentados adecuadamente con metadatos completos, sean accesibles en formatos abiertos y reutilizables, y cumplan con los estándares técnicos definidos por MinTIC. Este marco de calidad busca asegurar que los datos abiertos gubernamentales sean verdaderamente útiles para la ciudadanía, fomentando la transparencia y la participación ciudadana efectiva en línea con los principios del gobierno abierto.

Ilustración 5 Sello de Excelencia Gobierno Digital

Tomado de GUÍA DE ESTÁNDARES DE CALIDAD E INTEROPERABILIDAD DE LOS DATOS ABIERTOS DEL GOBIERNO DE COLOMBI. – MINTIC

8. PLAN DE CALIDAD DE DATOS

La planificación de la calidad de datos consiste en establecer una estrategia organizacional clara y estructurada, que permita asegurar que los datos y metadatos gestionados cumplan con los estándares de calidad requeridos. Este proceso implica definir roles y responsabilidades, desarrollar programas de calidad, y organizar las secuencias e interacciones de los procesos, además de implementar herramientas adecuadas para la gestión de calidad. Se busca garantizar que los datos sean confiables, accesibles y útiles para la toma de decisiones.

Planeación

Esta etapa consiste en evaluar el estado actual de los datos y metadatos gestionados por la organización. El diagnóstico permite identificar debilidades, inconsistencias y errores que afecten la calidad de los datos, tales como duplicados, datos incompletos, y falta de consistencia.

Con base en el diagnóstico, se define un plan estratégico que debe incluir objetivos claros, indicadores de desempeño y acciones específicas que se llevarán a cabo para mejorar la calidad de los datos. Este plan también debe estar alineado con las metas institucionales y normativas nacionales sobre gestión de datos.

Desarrollo

En esta fase, se calculan y planifican las intervenciones necesarias para mejorar la calidad de los datos. Esto puede incluir la implementación de nuevas herramientas o tecnologías, la capacitación del personal en mejores prácticas de gestión de datos, y la creación de procesos automatizados para la validación y limpieza de datos.

Para asegurar que las mejoras propuestas funcionen correctamente, se llevan a cabo pruebas técnicas. Las pruebas unitarias validan que cada componente del sistema de gestión de datos cumpla con su función, mientras que las pruebas funcionales aseguran que el sistema en su conjunto opera de acuerdo con los estándares de calidad.

Ejecución

Comunicación con la entidad mediante correos: La comunicación constante con los responsables de datos es crucial para mantenerlos informados sobre el estado del programa de calidad, así como de los resultados obtenidos y cualquier ajuste requerido. Esta comunicación puede realizarse mediante informes periódicos enviados por correo electrónico.

Seguimiento

Reunión periódica de seguimiento: Las reuniones de seguimiento permiten revisar los avances en la implementación del plan de calidad de datos, discutir los resultados obtenidos

y ajustar las estrategias según sea necesario. Estas reuniones deben ser regulares (mensuales o quincenales) y contar con la participación de todos los involucrados en la gestión de la calidad de los datos.

8.1. PLAN DE TRABAJO CALIDAD DE DATOS

Tabla 7 PLAN DE TRABAJO

Elaboración Propia

ACTIVIDAD	DESCRIPCIÓN	FECHA	RESPONSABLE	ENTREGABLE
Definición plan estratégico y objetivos	Establecimiento de política, KPIs y estrategias para mejorar la calidad de datos	I Y II Semestre	Dirección de Gobierno Digital	Política Y Plan de calidad de datos
Implementación de herramientas y tecnologías	Selección e implementación de herramientas de limpieza y validación de datos	III Semestre	Dirección de Gobierno Digital	Informe de implementación
Capacitación del personal	Generar estrategia con el fin de socializar en temas de calidad de datos	IV Y V Semestre	Dirección de Gobierno Digital	Reporte de calificaciones
Comunicación y coordinación con la entidad	Gestión de comunicaciones y alineación con diferentes Secretarías y dependencias	VI Semestre	Dirección de Gobierno Digital	Comunicaciones mediante cualquier medio como: correo electrónico y/o Forest
Reuniones de seguimiento	Monitoreo del avance y ajustes necesario	VII Y VIII Semestre	Dirección de Gobierno Digital	Mesas de trabajo, actas de reunión

La Dirección de Gobierno Digital coordina y supervisa la implementación y cumplimiento del Plan de Calidad de Datos de la Entidad. Todas las dependencias están obligadas a participar y cumplir con los requisitos establecidos en este documento.

El plan de trabajo tendrá una vigencia de cuatro años, por lo que las acciones a realizar están programadas en dicho periodo de tiempo (48 meses).

9. BIBLIOGRAFÍA

Ministerio de Tecnologías de la Información y las Comunicaciones. (s.f.). *Modelo de uso de calidad de datos*. Herramientas.datos.gov.co. Recuperado de https://herramientas.datos.gov.co/sites/default/files/Modelo_de_Uso_de_Calidad_de_Datos.pdf

Ministerio de Tecnologías de la Información y las Comunicaciones. (2022, julio 1). *Taller 2: Diagnóstico de calidad de datos*. Herramientas.datos.gov.co. Recuperado de https://herramientas.datos.gov.co/sites/default/files/2022-07/2022_07_01%20TALLER%202%20DIAG_CALIDAD.pdf

Ministerio de Tecnologías de la Información y las Comunicaciones. (2020, noviembre). *Guía de estándares de datos abiertos*. Herramientas.datos.gov.co. Recuperado de https://herramientas.datos.gov.co/sites/default/files/2020-11/A_guia_de_estandares_final_0.pdf

Ministerio de Tecnologías de la Información y las Comunicaciones. (s.f.). *Guía de datos abiertos de Colombia*. Herramientas.datos.gov.co. Recuperado de https://herramientas.datos.gov.co/sites/default/files/Guia%20de%20Datos%20Abiertos%20de%20Colombia.pdf

Ministerio de Tecnologías de la Información y las Comunicaciones. (s.f.). *Guía para la implementación de datos abiertos en Colombia*. Gobierno Digital. Recuperado de https://gobiernodigital.mintic.gov.co/692/articles-179118_recurso_5.pdf

Ministerio de Tecnologías de la Información y las Comunicaciones. (s.f.). *Marco de referencia para la arquitectura empresarial en entidades públicas de Colombia*. MinTIC. Recuperado de https://www.mintic.gov.co/arquitecturaempresarial/630/articles-237660_recurso_1.pdf

Ministerio de Tecnologías de la Información y las Comunicaciones. (s.f.). *Ciclo de vida de los datos: Anexo 1.2*. MinTIC. Recuperado de https://www.mintic.gov.co/portal/715/articles-198952_anexo_1_2_ciclo_vida_dato.pdf

Ministerio de Tecnologías de la Información y las Comunicaciones. (2021, agosto). *Guía de datos abiertos de Colombia*. Herramientas.datos.gov.co. Recuperado de https://herramientas.datos.gov.co/sites/default/files/2021-08/Guia%20de%20Datos%20Abiertos%20de%20Colombia.pdf