



TIC



DATOS ABIERTOS

Guía de estándares de **CALIDAD E INTEROPERABILIDAD** de los Datos Abiertos



Ministerio de Tecnologías de la Información y las Comunicaciones

Viceministerio de Transformación Digital

Dirección de Gobierno Digital

Julián Molina Gómez

Ministro de Tecnologías de la Información y las Comunicaciones

Yeimi Carina Murcia Yela

Viceministro de Transformación Digital

Lucy Urón Rincón

Director de Gobierno Digital

Natalia Albañil Riaño

Subdirectora de Fortalecimiento de Capacidades Digitales

Equipo Técnico Iniciativa Datos Abiertos

- Luisa Fernanda Medina, Asesora Viceministerio de Transformación Digital
Líder iniciativa nacional de datos abiertos
- Andrés Felipe Guerra, Consultor Técnico iniciativa Datos Abiertos
- Andrés Felipe Yanes, Consultor Lineamientos Datos Abiertos
- Luis Amílcar Bonivento, Consultor de Apropiación Iniciativa Datos Abiertos

CONTROL DE VERSIONES

Versión	Observaciones
Versión 1 Diciembre 2016	Guía para la apertura de datos en Colombia Dirigida a entidades del Estado para el desarrollo de procesos de apertura de datos públicos
Versión 2 Junio 2016	Guía de datos abiertos en Colombia Dirigida a las entidades sujeto de aplicación de la Ley 1712 de 2014 de Transparencia y Acceso a la Información Pública, para la aplicación de orientaciones y buenas prácticas en el desarrollo de estrategias de apertura y reúso de datos abiertos.
Versión 3 Junio 2016	Guía de datos abiertos en Colombia - modificaciones de diagramación
Versión 4 Enero 2019	Actualización temática de la Guía para el uso y aprovechamiento de Datos Abiertos en Colombia
Versión 5 Mayo 2019	Guía para el uso y aprovechamiento de Datos Abiertos en Colombia - modificaciones de diagramación
Versión 6 Septiembre 2019	Actualización proceso Ciclo de Datos alineado al reto de máxima velocidad.
Versión 7 Diciembre 2024	Actualización proceso con cambios en los parámetros de evaluación de calidad
Versión 8 Mayo 2025	Actualización de normatividad Conpes de Inteligencia Artificial 4144 de 2025

Comentarios, sugerencias o correcciones pueden ser enviadas al correo electrónico: datosabiertos@mintic.gov.co

Tabla de contenido

1. INTRODUCCIÓN.....	1
2. LA CALIDAD DE LOS DATOS	3
2.1. IMPORTANCIA DE PUBLICAR DATOS ABIERTOS DE CALIDAD	3
2.2. MEJORAMIENTO DE LA CALIDAD DE LOS DATOS	6
2.3. MARCO DE INTEROPERABILIDAD.....	8
3. CRITERIOS DE CALIDAD E INTEROPERABILIDAD	12
3.1. CRITERIO DE ACCESIBILIDAD	14
3.2. CRITERIO DE ACTUALIDAD	16
3.3. CRITERIO DE COMPLETITUD.....	19
3.4. CRITERIO DE COMPRESIBILIDAD	22
3.5. CRITERIO DE CONFORMIDAD	24
3.6. CRITERIO DE CONFIDENCIALIDAD	26
3.6.1. <i>Anonimizar los datos</i>	28
3.7. CRITERIO DE CONSISTENCIA.....	30
3.8. CRITERIO DE CREDIBILIDAD	32
3.9. CRITERIO DE DISPONIBILIDAD	34
3.10. CRITERIO DE EFICIENCIA	36
3.11. CRITERIO DE EXACTITUD	38
3.12. CRITERIO DE PORTABILIDAD.....	40
3.13. CRITERIO DE PRECISIÓN.....	41
3.14. CRITERIO DE RECUPERABILIDAD	43
3.15. CRITERIO DE RELEVANCIA.....	45
3.16. CRITERIO DE TRAZABILIDAD	47
3.17. CRITERIO DE UNICIDAD.....	50
4. LINEAMIENTOS PARA PREPARAR CONJUNTOS DE DATOS HABILITANTES PARA IA	52
4.1. ¿QUÉ ES UN CONJUNTO DE DATOS HABILITANTE PARA IA?.....	52
4.2. ¿CÓMO IDENTIFICAR O CONSTRUIR UN CONJUNTO DE DATOS HABILITANTE PARA IA EN UNA ENTIDAD PÚBLICA?	53
4.3. RECOMENDACIONES PARA LA PUBLICACIÓN DE CONJUNTOS DE DATOS ABIERTOS CON POTENCIAL PARA IA.....	55
4.4. METADATOS ENRIQUECIDOS Y ESTRUCTURACIÓN SEMÁNTICA PARA IA	57
4.5. EVALUACIÓN Y MEJORA CONTINUA DE LOS DATOS HABILITANTES PARA IA	59
5. GLOSARIO.....	61
6. REFERENCIAS.....	62

Lista de ilustraciones

ILUSTRACIÓN 1 ESTÁNDAR UNIVERSAL DE CALIDAD DE DOS CAPAS	5
ILUSTRACIÓN 2. CICLO DE VIDA DE LOS DATOS	7
ILUSTRACIÓN 4 MODELO CONCEPTUAL DEL MARCO DE INTEROPERABILIDAD	10
ILUSTRACIÓN 26 METADATOS ENLACE DE LA FUENTE	15
ILUSTRACIÓN 27 METADATOS URL DOCUMENTACIÓN Y URL NORMATIVA	16
ILUSTRACIÓN 10 EJEMPLO ÚLTIMA ACTUALIZACIÓN DE LOS DATOS	17
ILUSTRACIÓN 11 FRECUENCIA DE ACTUALIZACIÓN	18
ILUSTRACIÓN 19 CONJUNTO DE DATOS CON COLUMNAS VACÍAS	21
ILUSTRACIÓN 25 DESCRIPCIONES ADECUADAS PARA LAS COLUMNAS	23
ILUSTRACIÓN 15 REVISIÓN DE METADATOS PARA EL CRITERIO DE CONFORMIDAD	25
ILUSTRACIÓN 7 CALIFICACIÓN DE LA INFORMACIÓN DE ACUERDO CON SUS NIVELES DE SEGURIDAD	27
ILUSTRACIÓN 20 CONJUNTO DE DATOS CON DATOS INCONSISTENTES	32
ILUSTRACIÓN 24 METADATOS CREDIBILIDAD	34
ILUSTRACIÓN 29 ERROR EN EFICIENCIA.....	37
ILUSTRACIÓN 16 REVISIÓN DE METADATOS PARA EL CRITERIO DE EXACTITUD SINTÁCTICA.....	39
ILUSTRACIÓN 22 FORMATOS PARA LA PUBLICACIÓN DE DATOS ABIERTOS	41
ILUSTRACIÓN 21 EJEMPLO ERROR PRECISIÓN	43
ILUSTRACIÓN 9 MUESTRA DE CONJUNTO DE DATOS CON INFORMACIÓN NO RELEVANTE	47
ILUSTRACIÓN 28 EVIDENCIA DE FILAS Y COLUMNAS DUPLICADAS	51

1. Introducción

En los últimos años, Colombia ha avanzado en la consolidación de un ecosistema de datos abiertos robusto y funcional, articulado a través de la plataforma nacional www.datos.gov.co. Esta plataforma permite a las entidades públicas del orden nacional y territorial publicar información de interés general en formatos accesibles, reutilizables y abiertos, como parte del cumplimiento del derecho fundamental de acceso a la información pública y de los principios de transparencia, participación e innovación digital.

No obstante, la publicación de datos por sí sola no garantiza su aprovechamiento. Para que los datos abiertos puedan ser utilizados de forma efectiva por ciudadanos, empresas, academia y por las propias entidades del Estado, deben cumplir condiciones técnicas y semánticas que aseguren su calidad, interoperabilidad y trazabilidad. Estos atributos son fundamentales para facilitar su integración en sistemas de información, su vinculación con modelos de análisis y su reutilización en soluciones tecnológicas, incluyendo aplicaciones basadas en inteligencia artificial (IA).

La presente guía tiene como objetivo brindar orientaciones prácticas a los líderes de tecnología y responsables de datos en las entidades públicas del orden nacional y territorial, para asegurar que los conjuntos de datos abiertos cumplan con estándares básicos de calidad e interoperabilidad. Se enfoca en ofrecer un marco claro para planificar, evaluar y mejorar la publicación de datos, alineado con políticas públicas nacionales, la Hoja de Ruta Nacional y Sectoriales de Datos Abiertos Estratégicos, y referentes internacionales como la Carta Internacional de Datos Abiertos y la norma ISO/IEC 25012.

Este documento proporciona criterios de calidad comprensibles y aplicables en el contexto institucional, con el fin de fortalecer la capacidad de las entidades para publicar datos útiles, confiables y sostenibles en el tiempo.

Esta actualización se enmarca además en el CONPES 4144 de 2025 sobre Inteligencia Artificial, particularmente en el compromiso 2.20, que establece la necesidad de actualizar los lineamientos de datos abiertos como insumo estratégico para el desarrollo responsable de sistemas de IA. Por tanto, esta guía no solo orienta sobre buenas prácticas en calidad de datos, sino que también sienta las bases para su



aprovechamiento por parte de soluciones tecnológicas avanzadas, promoviendo el uso ético, seguro y soberano de los datos en entornos digitales e inteligentes.

2. La calidad de los datos

2.1. Importancia de publicar datos abiertos de calidad

En la actualidad, la publicación de datos abiertos no se limita a liberar archivos en línea. Para que esta práctica genere valor público, es indispensable que los datos sean útiles, confiables y reutilizables, lo cual solo se logra si cumplen con estándares mínimos de calidad.

Desde la década de 1950, el concepto de calidad ha evolucionado desde su aplicación a productos físicos hacia bienes intangibles como los datos y el software. Diversos expertos y normas internacionales han aportado definiciones que hoy son base para el trabajo en entidades públicas:

- **Philip Crosby¹**: Calidad es cumplir con los requisitos definidos.
- **Wang & Strong²** : Calidad es que los datos sean aptos para el uso que se espera de ellos.
- **ISO 9001:2000³**: Calidad es el grado en que las características de un producto o servicio cumplen con los requisitos establecidos.
- **ISO 9001:2008 y 2015⁴**: La calidad incluye también la sostenibilidad, la gestión de riesgos y las expectativas de las partes interesadas

En el contexto de la gestión pública, esto significa que los datos deben estar preparados para apoyar decisiones, articular políticas, fomentar innovación y facilitar el control ciudadano.

¹ https://es.wikipedia.org/wiki/Philip_Crosby

² <https://www.chospab.es/calidad/archivos/Documentos/NormalInternacionalISO9001.pdf>

³ <https://www.iso.org/obp/ui/#iso:std:iso:9001:ed-4:v2:es>

⁴ <https://www.iso.org/obp/ui/#iso:std:iso:9001:ed-5:v1:es>

Con la entrada en vigor del CONPES 4144 de 2025 sobre Inteligencia Artificial, se refuerza la necesidad de contar con datos abiertos de calidad como insumo estratégico para el desarrollo y despliegue de sistemas inteligentes confiables, éticos y auditables. La calidad de los datos abiertos no solo incide en su reutilización básica, sino que determina la viabilidad técnica, normativa y social de la inteligencia artificial en los servicios públicos.

Normas técnicas que respaldan el concepto de calidad

Este modelo, reafirmado en la norma **ISO/IEC 25012** propone un modelo de calidad de datos que distingue dos dimensiones clave⁵

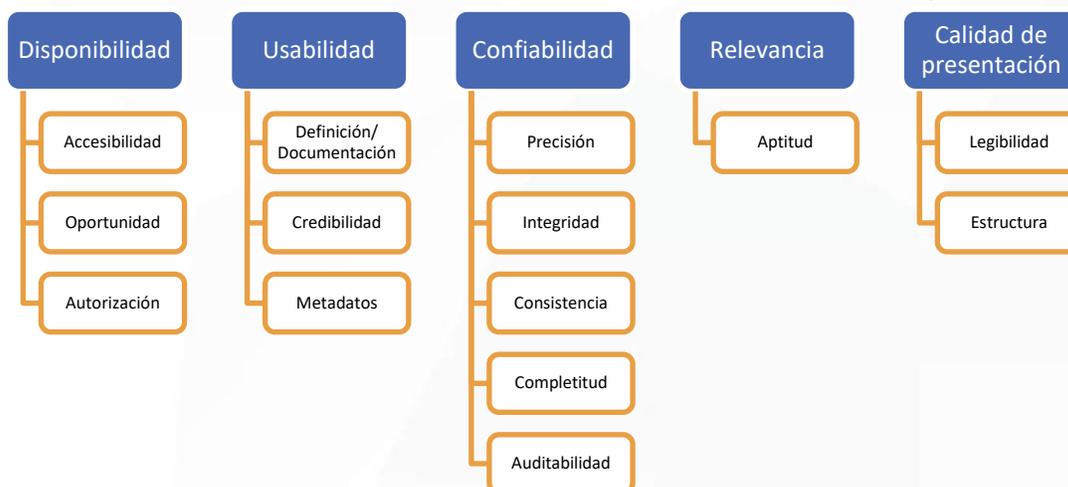
- **Calidad inherente de los datos:** Qué tan bien los datos, por sí mismos, pueden cumplir su propósito.
- **Calidad dependiente del sistema:** Qué tanto el entorno tecnológico permite mantener y aprovechar esa calidad.

Este modelo, reafirmado en la norma **UNE-ISO/IEC 25012:2023**⁶, ha sido adoptado por gobiernos y organismos multilaterales como guía para evaluar y mejorar la publicación de datos.

⁵ <https://iso25000.com/index.php/normas-iso-25000/iso-25012>

⁶ <https://www.une.org/encuentra-tu-norma/busca-tu-norma/norma?c=N0071618>

Ilustración 1 Estándar universal de calidad de dos capas



Fuente. Tomado de Los desafíos de la calidad de los datos y la evaluación de la calidad de los datos en la era del Big Data, por Cai. L. & Yangyon Z., 2015, Data Science Journal 8 (Cai, 2015)

Dimensiones clave de la calidad de los datos

De forma práctica, las siguientes dimensiones son fundamentales para asegurar que los datos abiertos cumplan su propósito:

- **Precisión:** Representan fielmente la realidad.
- **Integridad:** Medida en que los datos están completos y sin valores faltantes.
- **Validez:** Cumplen con formatos y reglas esperadas.
- **Consistencia:** Son coherentes entre distintas fuentes y periodos.
- **Unicidad:** Ausencia de duplicados en los datos.
- **Actualización:** Están vigentes para su uso.
- **Relevancia:** Son útiles para el análisis y la toma de decisiones⁷

Beneficios de publicar datos con calidad

Contar con datos abiertos de calidad trae beneficios claros para las entidades públicas y sus grupos de interés:

⁷ IBM Research (s.f.). Data Quality Dimensions.

- **Mayor valor público:** La ciudadanía, la academia y las empresas pueden reutilizar los datos para crear soluciones, productos y análisis.
- **Mejor interoperabilidad:** Se reducen redundancias y se facilita el intercambio eficiente de información entre entidades.
- **Impulso a la innovación:** Tecnologías como inteligencia artificial requieren datos confiables para ofrecer resultados precisos.
- **Transparencia y legitimidad:** Datos claros y accesibles fortalecen la confianza en las instituciones.

Importancia para los líderes TIC

Para los CIOs, líderes de TI y responsables de datos, la calidad no debe verse como una tarea técnica adicional, sino como una estrategia organizacional que potencia el valor de la información que ya se produce en la entidad. Esta guía busca facilitar ese camino, proponiendo criterios claros, prácticos y adaptables para fortalecer la calidad e interoperabilidad de los datos abiertos en Colombia. En el marco del CONPES 4144, esta labor es además clave para que las entidades puedan habilitar el uso responsable y ético de la IA, garantizando que los datos que nutren los modelos sean confiables, auditables y alineados con las necesidades del país.

2.2. Mejoramiento de la calidad de los datos

El mejoramiento continuo de la calidad de los datos es un proceso estratégico que requiere compromiso institucional, capacidades técnicas y liderazgo. Para los líderes de TI (CIOs) y responsables de datos, este reto implica no solo dar cumplimiento a los lineamientos normativos, sino también impulsar el valor de la información como un activo clave para la toma de decisiones, la mejora de servicios y la innovación digital en el sector público.

Calidad como proceso continuo en la gestión pública

La Guía para el Uso y Aprovechamiento de Datos Abiertos en Colombia⁸ resalta que una estrategia de apertura de datos trasciende la simple preparación y publicación de información por parte de las entidades públicas. Esta estrategia exige acciones concretas orientadas a fomentar el uso efectivo de los datos publicados, realizar un seguimiento riguroso y evaluar el valor agregado que generan en su aplicación.

En este marco, la segunda fase del proceso de apertura, denominada “**Monitoreo de la calidad y el uso**”, es un componente esencial para consolidar la utilidad de los datos abiertos. Esta guía profundiza en esa etapa, ofreciendo un enfoque práctico y aplicable, orientado a que las entidades puedan cumplir con criterios estandarizados que fortalezcan la calidad, la sostenibilidad y la interoperabilidad de sus conjuntos de datos.

Ilustración 2. Ciclo de vida de los datos



Fuente. Guía de uso y aprovechamiento de datos en Colombia

⁸ https://drive.google.com/file/d/1jiuxTrsXyz_rEtkTpG9MR3qC-yTGUhw_/view

Rol de los CIOs y responsables de datos

Para lograr mejoras sostenibles en la calidad de los datos abiertos, los CIOs y líderes de TI deben:

- Identificar las brechas de calidad actuales en los datos ya publicados.
- Fortalecer los procesos institucionales de planeación, documentación, actualización y validación de datos.
- Asegurar la trazabilidad y auditabilidad de los datos, especialmente cuando alimentan procesos de automatización o análisis con IA.
- Promover la formación de equipos interdisciplinarios, en donde los líderes TIC trabajen articuladamente con las áreas misionales y de planeación para mejorar la calidad de los datos desde su origen.
- Incorporar herramientas y estándares técnicos compatibles con los principios de interoperabilidad, soberanía tecnológica y ética en el uso de datos, tal como lo plantea el marco de IA responsable en Colombia.

📌 *Nota práctica:* La guía nacional de calidad recomienda implementar un enfoque gradual, iniciando con un diagnóstico básico de los conjuntos de datos más consultados y priorizando acciones de mejora según su valor estratégico y nivel de reutilización potencial.

2.3. Marco de interoperabilidad

¿Por qué es clave la interoperabilidad en el contexto de los datos abiertos?

Publicar datos abiertos no solo implica disponibilizarlos en línea, sino también garantizar que puedan ser comprendidos, conectados e integrados por otras entidades públicas, empresas y ciudadanos. Para lograrlo, es esencial que los datos sigan principios de interoperabilidad, es decir, que puedan ser intercambiados y reutilizados entre sistemas, personas y organizaciones sin fricciones técnicas o semánticas.

La interoperabilidad promueve una visión unificada del Estado, donde las entidades trabajan de manera articulada para ofrecer servicios digitales más

eficientes, confiables y centrados en el ciudadano. En este sentido, el Marco de Interoperabilidad del MinTIC es la referencia nacional para facilitar el intercambio de información entre entidades públicas, bajo cuatro dominios fundamentales: político-legal, organizacional, semántico y técnico

Objetivos del Marco de Interoperabilidad

Este marco busca:

- **Facilitar la prestación de servicios públicos digitales** basados en datos reutilizables por defecto.
- **Establecer condiciones comunes para el intercambio de información** entre entidades.
- **Promover una visión unificada del Estado**, apoyada en la colaboración digital entre instituciones.
- **Contribuir al uso estratégico de los datos abiertos**, asegurando su integración y aprovechamiento en todo el ciclo de vida de la información.

Aplicación en la calidad de los datos abiertos

La publicación de datos abiertos debe considerar los lineamientos del marco para garantizar que estos:

- Estén alineados con **estándares técnicos y semánticos comunes**.
- Sean interpretables por múltiples actores del Estado.
- Se encuentren disponibles para ser **compartidos, integrados y reutilizados**, sin barreras técnicas o legales.

• Ejemplo: Cuando una entidad publica un conjunto de datos sobre contrataciones públicas, debe asegurarse de que los campos como "modalidad de contratación" o "estado del contrato" estén definidos según el **Lenguaje Común de Intercambio de Información** del MinTIC, permitiendo que otras entidades los interpreten de forma uniforme.

Cuatro dominios de interoperabilidad para la gestión de datos abiertos

Los CIOs deben comprender que la interoperabilidad va más allá del aspecto técnico. El marco se estructura en los siguientes dominios:

1. **Dominio Político-Legal:** Asegura que el intercambio de información se ajuste al marco jurídico vigente. Las entidades deben revisar y aplicar instrumentos

legales (como acuerdos interinstitucionales o decretos) que permitan compartir datos sin afectar la protección de datos personales.

2. **Dominio Organizacional:** Se refiere a los procesos, roles y estructuras internas necesarias para garantizar que los servicios de datos abiertos estén alineados con los objetivos misionales de la entidad y que existan responsables claros del flujo y la calidad de la información.
3. **Dominio Semántico:** Asegura que los datos tengan un significado común entre entidades. Esto implica usar el Lenguaje Común de Intercambio de Información, publicado por el MinTIC, para estructurar y definir datos de forma estandarizada.
4. **Dominio Técnico:** Se enfoca en los mecanismos tecnológicos que habilitan el intercambio digital. Aunque esta guía no aborda aspectos técnicos, es clave que los CIOs garanticen la existencia de canales seguros y eficientes para compartir datos con otras entidades.

Ilustración 3 Modelo conceptual del Marco de Interoperabilidad



Fuente. Ministerio TIC.

✎ Esta guía complementa estos dominios enfocándose especialmente en los aspectos semánticos y técnicos que inciden directamente en la calidad y reutilización de los datos abiertos.

El Marco también incorpora un modelo de madurez, que permite a las entidades evaluar su progreso en la adopción de prácticas interoperables. Este modelo puede usarse para priorizar mejoras en la gestión de datos abiertos, asignar recursos y establecer metas concretas en interoperabilidad.

La gobernanza de la interoperabilidad implica definir responsables dentro de la entidad, formalizar acuerdos de intercambio y asegurar la calidad de los datos compartidos. Esto debe incorporarse al Plan Estratégico de Tecnologías de la Información (PETI) y a los lineamientos de transformación digital institucional.

Recomendaciones prácticas para entidades públicas

- Adopta los lineamientos del Marco de Interoperabilidad como referencia institucional, en especial al planear la publicación de conjuntos de datos abiertos.
- Utiliza el Lenguaje Común del MinTIC para describir y estandarizar los elementos de dato incluidos en tus conjuntos de datos.
- Revisa si tu entidad cuenta con acuerdos interinstitucionales vigentes para el intercambio de información, y ajusta su uso conforme a la ley.
- Evalúa el nivel de madurez de tu entidad usando la herramienta publicada por el MinTIC, e implementa mejoras progresivas.

Interoperabilidad como parte del ciclo de vida de los datos

Publicar datos abiertos de calidad requiere planear su interoperabilidad desde el inicio. Las entidades deben considerar:

1. Revisión normativa sobre el uso y publicación del dato.
2. Estandarización semántica (glosarios, catálogos de datos, lenguaje común).
3. Validación técnica de estructura y formato interoperable.
4. Documentación clara y completa (metadatos y licencias).
5. Monitoreo del uso e intercambio para retroalimentar mejoras.



Puedes consultar este estándar en el portal oficial:

<http://lenguaje.mintic.gov.co>

Adoptar el Marco de Interoperabilidad no es una tarea técnica aislada. Es una decisión estratégica que permite a las entidades públicas cumplir con los principios de transparencia, eficiencia e innovación mediante el uso inteligente y articulado de la información. En el contexto de los datos abiertos, este marco es una herramienta indispensable para asegurar que la información publicada tenga calidad, utilidad y un verdadero impacto en la ciudadanía.

3. Criterios de Calidad e Interoperabilidad

Para que los datos abiertos realmente generen valor público, deben cumplir con criterios básicos de **calidad** e **interoperabilidad**, que permitan su comprensión, reutilización y vinculación con otros conjuntos de datos. Estos criterios funcionan como una guía práctica para que las entidades públicas evalúen, mejoren y mantengan la integridad, utilidad y accesibilidad de la información publicada.

A continuación, se presentan los principales criterios que deben ser considerados por las entidades antes, durante y después de publicar conjuntos de datos en el portal nacional datos.gov.co.

#	Criterio	Descripción	Aplicación práctica
---	----------	-------------	---------------------

1	Accesibilidad	El conjunto puede ser consultado y descargado por quien lo necesite.	Sin requisitos de registro, contraseña ni software especial para acceder.
2	Actualidad	Vigencia y actualización de los datos publicados.	Los datos reflejan el estado más reciente según su periodicidad declarada.
3	Compleitud	Todos los campos obligatorios están diligenciados.	Ningún campo crítico queda en blanco (Ej: NIT, fecha, valor).
4	Comprensibilidad	Los datos pueden ser interpretados fácilmente por cualquier usuario.	Encabezados claros, glosarios, campos con nombres comprensibles.
5	Conformidad	Cumplimiento de lineamientos y estándares vigentes.	Uso de plantillas de MinTIC, estándares abiertos y normativas institucionales.
6	Confidencialidad	Los datos solo deben ser accedidos por personal autorizado.	Accesos restringidos en sistemas internos; no publicar datos personales sin anonimizar.
7	Consistencia	Datos coherentes y sin contradicción.	Fechas no desordenadas, estados válidos, formatos coherentes en todos los registros.
8	Credibilidad	Información veraz y confiable para los usuarios.	Fuente oficial declarada y responsable institucional visible.
9	Disponibilidad	Los datos están en línea cuando se necesitan.	El enlace al conjunto funciona 24/7 y no presenta errores o caídas frecuentes.
10	Eficiencia	Plataforma permite análisis y descargas con buen rendimiento.	Portal rápido, sin errores al descargar o consultar.
11	Exactitud	Datos diligenciados correctamente.	No hay errores de escritura, ortografía o campos mal llenados.
12	Portabilidad	Formatos sin restricciones para su reutilización.	Publicación en CSV, JSON o Excel limpio, sin bloqueos ni macros.
13	Precisión	Nivel de desagregación de los datos es adecuado al original.	Datos publicados a nivel municipio si así fueron recolectados, no agregados por región.

14	Recuperabilidad	Capacidad de restaurar o recuperar datos si se pierden o fallan.	Copias de seguridad del dataset y control de versiones.
15	Relevancia	Los datos publicados deben ser de utilidad para los usuarios.	Conjuntos alineados con demandas ciudadanas o estratégicas del sector.
16	Trazabilidad	Histórico del conjunto de datos: fechas de creación, publicación y actualizaciones.	Registro de versiones del dataset y fechas en metadatos.
17	Unicidad	Detección de registros duplicados e identificación única.	Campos como NIT, ID o número de contrato permiten identificar unívocamente registros.

3.1. Criterio de accesibilidad

Definición

El criterio de **accesibilidad** evalúa la facilidad con la que los usuarios pueden **encontrar, consultar y utilizar un conjunto de datos abiertos**. Se refiere no solo a la disponibilidad técnica de los datos, sino también a las condiciones para su acceso libre, sin restricciones administrativas ni barreras tecnológicas.

La accesibilidad implica que los datos estén disponibles en formatos abiertos, acompañados de metadatos descriptivos (etiquetas y enlaces), y que puedan ser accedidos de manera gratuita y sin requerimientos de registro o autenticación.

Ilustración 4 Metadatos enlace de la fuente



Fuente. Captura de pantalla de datos.gov.co

⚙️ Aplicación práctica

Una entidad pública cumple con este criterio cuando:

- Publica el conjunto en un **portal oficial, centralizado y accesible** como datos.gov.co.
- Incluye **etiquetas o palabras clave** que facilitan la búsqueda y clasificación del conjunto.
- Proporciona **vínculos de atribución o enlaces relevantes**, como:
 - Enlace a la fuente original.
 - URL de la documentación técnica del conjunto.
 - URL de la normativa o acto administrativo que respalda el conjunto.
- No exige **registro ni autenticación** para consultar o descargar los datos.
- Publica los datos bajo una **licencia abierta**, sin restricciones de uso.

💡 Ejemplo aplicado

Caso: Una entidad ambiental publica un conjunto de datos sobre calidad del aire.

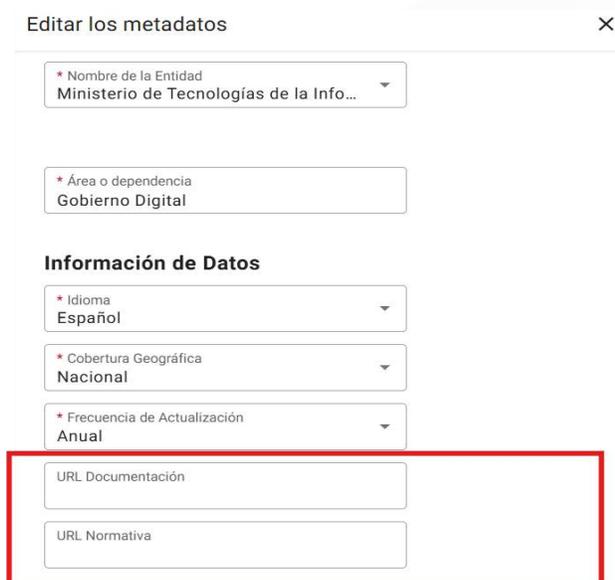
Acciones que fortalecen la accesibilidad:

- Agrega etiquetas como “medio ambiente”, “contaminación”, “aire”.
- Incluye una URL al sistema de monitoreo en tiempo real.
- Enlaza la resolución que obliga a publicar dicha información.
- El conjunto está disponible en CSV y se puede descargar sin necesidad de registrarse.

La siguiente imagen muestran cómo diligenciar correctamente los campos de

metadatos en datos.gov.co para cumplir con este criterio.

Ilustración 5 Metadatos URL documentación y URL normativa



Editar los metadatos

* Nombre de la Entidad
Ministerio de Tecnologías de la Info...

* Área o dependencia
Gobierno Digital

Información de Datos

* Idioma
Español

* Cobertura Geográfica
Nacional

* Frecuencia de Actualización
Anual

URL Documentación

URL Normativa

Fuente. Captura de pantalla de datos.gov.co

✓ Buenas prácticas institucionales

- Incluir **al menos una etiqueta y un vínculo de referencia** en todos los conjuntos.
- Validar que los enlaces funcionen correctamente y lleven a sitios oficiales.
- Mantener consistencia en el uso de etiquetas para facilitar búsquedas temáticas.
- Asegurarse de que los datos estén disponibles sin pasos innecesarios para el usuario final.

3.2. Criterio de actualidad

✚ Definición

El criterio de **actualidad** evalúa si los datos publicados están **vigentes y alineados con la periodicidad definida** para su actualización. Este criterio es clave para asegurar que los conjuntos de datos reflejen la realidad en el momento de su consulta y, por

tanto, puedan ser utilizados como base confiable para la toma de decisiones públicas, investigaciones o desarrollo de servicios.

La actualidad se mide a partir de dos elementos clave en los metadatos:

- **Fecha de última actualización del conjunto.**
- **Frecuencia de actualización esperada**, según el tipo de información.

Ilustración 6 Ejemplo última actualización de los datos

Actualizado	
1 de abril de 2024	
Última actualización de los datos	Última actualización de metadatos
23 de septiembre de 2022	1 de abril de 2024
Fecha de creación	
10 de julio de 2018	

Fuente. Captura de pantalla de datos.gov.co

Ilustración 7 Frecuencia de Actualización

Editar los metadatos ×

* Nombre de la Entidad
Ministerio de Tecnologías de la Info...

* Área o dependencia
Gobierno Digital

Información de Datos

* Idioma
Español

* Cobertura Geográfica
Nacional

* Frecuencia de Actualización
Anual

URL Documentación

URL Normativa

Fuente. Captura de pantalla de datos.gov.co

⚙️ Aplicación práctica

Una entidad pública cumple con el criterio de actualidad cuando:

- La **última fecha de actualización** del conjunto es reciente y coherente con la frecuencia declarada.
- Se ha especificado correctamente la **frecuencia de actualización** en los metadatos (mensual, trimestral, anual, etc.).
- Se mantiene un control interno para actualizar el conjunto de datos en los plazos definidos.
- Los datos reflejan el ciclo real de producción o captura de la información.

💡 Ejemplo aplicado

Caso: Una entidad del sector transporte publica datos sobre accidentes viales.

Acciones recomendadas:

- Define en los metadatos que la frecuencia de actualización es **mensual**.
- Actualiza efectivamente el conjunto cada mes, registrando la **fecha exacta de la última actualización**.
- Si no hay nueva información en un mes determinado, incluye una nota aclaratoria en los metadatos.

✅ **Buenas prácticas institucionales**

- Alinear la frecuencia de publicación con la **disponibilidad real de los datos** en la entidad.
- Establecer alertas o recordatorios internos para cumplir con los ciclos de actualización.
- No cambiar la fecha de actualización si no se han modificado los datos efectivamente.
- Registrar en los metadatos si un conjunto está **archivado**, **cerrado** o sin nuevos registros por razones justificadas.

3.3. Criterio de completitud

✳ **Definición**

El criterio de **completitud** mide qué tan **completos y útiles** son los datos publicados para cumplir con su propósito de análisis, consulta o reutilización. Evalúa la **proporción de valores diligenciados** y la **presencia de todas las columnas necesarias**, según la estructura definida por la entidad y descrita en los metadatos.

La incompletitud de un conjunto de datos reduce su calidad, genera ambigüedad y puede limitar su aprovechamiento por parte de la ciudadanía, el sector privado o los organismos de control.

⚙️ Aplicación práctica

Una entidad pública cumple con este criterio cuando:

- La mayoría de los **campos esperados** están diligenciados, sin celdas vacías innecesarias.
- No existen **columnas vacías o sin datos relevantes**.
- Se han incluido etiquetas como “No aplica”, “No disponible” o “Otro” en campos de texto, evitando espacios vacíos.
- Se consolidan los datos relacionados en un solo conjunto, en lugar de publicarlos fragmentados o dispersos.

✦ **Importante:** Publicar un conjunto con una sola columna, o con múltiples columnas sin valores, se considera una mala práctica y afecta negativamente la calidad percibida.

Para obtener un puntaje alto en este criterio, es importante incluir opciones como “No aplica”, “Otro”, “No sabe”, “No responde” o “No disponible” en los diccionarios de datos o etiquetas de campo de tipo texto. Esto ayuda a evitar espacios vacíos cuando no se cuenta con la información necesaria. Además, se recomienda evitar publicar conjuntos de datos con una sola columna o que estén incompletos. Si existen datos relacionados con un mismo tema, pero están distribuidos en diferentes fuentes, es mejor consolidarlos en un único conjunto de datos que sea lo más completo posible para facilitar su uso por los usuarios.

💡 Ejemplo aplicado

Caso: Una entidad pública desea publicar datos de vigilancia epidemiológica con registros por municipio.

Situación inicial:

- El conjunto contiene columnas para municipio, fecha, tipo de enfermedad, y número de casos.
- Algunas columnas presentan celdas vacías o sin valores diligenciados.

Acciones recomendadas:

- Completar los campos vacíos con valores tipo “No aplica” o “Sin reporte”.
- Eliminar columnas vacías si no se tiene información para publicar.

- Validar contra el diccionario de datos que todas las columnas esperadas estén incluidas.

La siguiente imagen muestra cómo la ausencia de información diligenciada afecta negativamente el cumplimiento de este criterio.

Ilustración 8 Conjunto de datos con columnas vacías

Categoría de la In...	Nombre de la Inf...	Descripción	Área Responsable	Dueño de la Infor...	Responsable de I...	Idioma	Clasificación de I...	Fundamento Legal
Intranet	Información (Datos)	Office (Word-Excel,PP)	Extranet		Generar	Pública	Asesor de Presidencia	Presidente
Correo Electrónico	Sistema de Informació	BD	Intranet		Almacenar	Pública Clasificada	Asistente Administrati	Secretario General
Archivo Físico	Infraestructura	Correo Electrónico	Archivo Físico		Procesar	Pública Reservada	Auxiliar de Oficina Gra	Vicepresidente de Oj
PC	Instalaciones	PDF	Correo Electrónico		Consultar		Conductor de Preside	Vicepresidente Técni
Sist de Info (SISE - SIA	Recurso Humano	Archivo Físico	Carpeta Compartida		Recibir		Gerente	Vicepresidente Adm
Log de Auditoria		Cintas	Medios de Alm Ext		Publicar		Gerente de Área Grad	Vicepresidente de In
Medios de alma Ext		Imagen			Archivar		Gerente Sucursal Tipo	Vicepresidente de Py
Recurso Compartido		Audios			Enviar		Gerente Sucursal Tipo	Jefe de la OGR
BD Misionales		Videos			Verificar		Gerente Sucursal Tipo	Jefe de la OTI

Fuente. Captura de pantalla de datos.gov.co

✓ Buenas prácticas institucionales

- Aplicar controles de calidad antes de la publicación, para detectar campos o columnas sin información.
- Completar los campos con valores estándar (“No disponible”, “No aplica”, etc.) en lugar de dejarlos en blanco.
- Consolidar los datos en un solo conjunto siempre que sea posible, evitando dispersión temática.
- Validar que el conjunto esté alineado con los metadatos y el diccionario de datos declarado.

3.4. Criterio de Comprensibilidad

Definición

El criterio de **comprensibilidad** evalúa qué tan claro, legible y coherente es un conjunto de datos desde la perspectiva de los usuarios. Se refiere a la forma en que están estructurados los **metadatos, nombres de campos, descripciones** y otros elementos que permiten interpretar adecuadamente el contenido del conjunto de datos.

La comprensibilidad no depende solo de que la información esté presente, sino de que esté escrita con un lenguaje claro, sin ambigüedades, y sea útil para diferentes tipos de usuarios, desde ciudadanos hasta analistas técnicos.

Aplicación práctica

Una entidad pública cumple con este criterio cuando:

- El **título** del conjunto de datos es claro y refleja fielmente su contenido.
- La **descripción general** del conjunto explica adecuadamente su contexto, incluye definiciones clave y es comprensible sin necesidad de conocimientos especializados.
- Cada **campo o columna** tiene un nombre descriptivo y una explicación mínima de 15 caracteres en los metadatos.
- Se evita el uso de códigos crípticos, abreviaturas ambiguas o nombres genéricos como “dato1”, “columnaA”.
- Si se usan códigos (como P1, P2, etc.), se acompañan de una definición clara en los metadatos o un diccionario de datos.

Ejemplo aplicado

Caso: Una entidad publica un conjunto de datos sobre resultados de encuestas institucionales.

Situación inicial:

- Los campos aparecen como “P1”, “P2”, “P3”, sin ninguna explicación adicional.
- La descripción del conjunto solo dice “Encuesta anual”.

Acciones recomendadas:

- Ampliar el título a: “Resultados encuesta de percepción institucional – Año 2024”.
- Incluir una descripción general explicando el propósito, población encuestada y método.
- Describir cada pregunta (P1: “Satisfacción con el servicio recibido”, etc.).
- Corregir errores ortográficos y evitar repeticiones entre el nombre del campo y su descripción.

Esta imagen muestra cómo deben aparecer los campos de descripción en los metadatos de columnas, facilitando su comprensión por parte de cualquier usuario.

Ilustración 9 Descripciones adecuadas para las columnas

 Fecha de Generación de la información (AAAA/MM/DD)	Fecha en que se genera la información	fecha_de_generaci_n_de_la	Marca de tiempo variable
 Nombre del responsable producción de la información (Propietario)	Entidad responsable de producir la información	nombre_del_responsable	Texto
 Nombre del responsable de la información (Custodio)	Entidad responsable de custodiar la información	nombre_del_responsable_de	Texto
 Objetivo legítimo de la excepción	Norma que legitima la excepción	objetivo_leg_timo_de_la	Texto
 Fundamento constitucional o legal	Norma constitucional	fundamento_constitucional	Texto
 Fundamento jurídico de la excepción	Norma jurídica	fundamento_jur_dico_de_la	Texto
 Excepción Total o Parcial	Si la excepción es total o parcial	excepci_n_total_o_parcial	Texto
 Fecha de la calificación (AAAA/MM/DD)	Fecha de clasificación del activo	fecha_de_la_calificaci_n	Marca de tiempo variable

Fuente. Captura de pantalla de datos.gov.co

Nota: Asegúrese de que las siglas utilizadas sean claras y fáciles de entender, y revise que no existan errores ortográficos, ya que estos pueden afectar la comprensión y credibilidad del conjunto de datos.

Buenas prácticas institucionales

- Evitar repetir el nombre de la columna como descripción (ejemplo incorrecto: columna “Región” con descripción “Región”).
- Utilizar un lenguaje claro, sin tecnicismos innecesarios.
- Revisar ortografía, redacción y coherencia semántica.

- Usar glosarios o descripciones ampliadas cuando se utilicen clasificaciones complejas.
- Verificar que la descripción de cada campo tenga **mínimo 15 caracteres útiles** y aporte valor interpretativo.

3.5. Criterio de conformidad

Definición

El criterio de **conformidad** evalúa en qué medida los valores dentro de un conjunto de datos **cumplen con las normas, formatos, estándares y convenciones definidas** por la entidad o por marcos de referencia nacionales. Esto incluye la validación contra listas oficiales (como departamentos, municipios, años, coordenadas geográficas) y el uso correcto del tipo de datos esperado (numérico, texto, fecha, etc.).

La conformidad es fundamental para garantizar que los datos sean **coherentes, interoperables y confiables**, tanto dentro de una entidad como en procesos de intercambio entre instituciones.

Aplicación práctica

Una entidad pública cumple con este criterio cuando:

- Los campos contienen **valores válidos** conforme a listas de referencia oficiales o catálogos estandarizados (ej. DANE, MinTIC).
- Los tipos de datos en cada campo coinciden con los definidos en los metadatos (ej. no hay letras en campos numéricos).
- No se incluyen **valores ficticios o de relleno** como “1234”, “XXXXX” o “N/A” sin justificación.
- Se utiliza lógica de validación para detectar errores en nombres de municipios, siglas mal escritas o fechas inválidas.
- Las **coordenadas geográficas** tienen formato correcto (latitud/longitud) y están dentro del rango permitido.

Ejemplo aplicado

Caso: Una entidad publica un conjunto con datos de ubicación de puntos de atención en salud.

Situación problemática:

- En el campo “Departamento”, aparecen valores como “Tolimaa”, “Atlantico” (sin tilde), y “Desconocido”.
- En el campo “Latitud”, se incluyen valores vacíos o letras.

Acciones recomendadas:

- Validar el campo “Departamento” contra la lista oficial de los 32 departamentos de Colombia.
- Asegurar que las coordenadas estén completas y en el formato decimal correcto.
- Establecer controles internos para evitar valores no válidos antes de publicar el conjunto.

Esta imagen muestra cómo el portal de datos abiertos permite identificar metadatos incompletos o valores fuera de estándar, penalizando el cumplimiento del criterio del conjunto.

Ilustración 10 Revisión de metadatos para el criterio de conformidad

Departamento	Municipio	Año	Latitud	Longitud	Censo Electrónico
Antioquia	Medellín	2020	6.2442	-75.5812	juan@example.com
Bogotá		2021	4.711	-74.0721	
Desconocido	Bogotá				maria@
Cundinamarca	Sococha	Desconocido	4.589	-74.083	sofia@
Error	Error	2023	Invalid	Desconocido	
Atlantico	Barranquilla	XXXX			carlos@example
Bolívar	Cartagena	2019	10.391	-75.482	pedro@example.com
Invalid	Popayán	2025	2.444	-76.606	error@
Cauca	Pasto		Error	Error	
	Desconocido	2021	-1.208	-77.281	luna@

Fuente. Captura de pantalla de datos.gov.co

✓ Buenas prácticas institucionales

- Establecer catálogos oficiales de referencia para campos como: departamentos, municipios, tipos de contrato, sectores económicos, etc.

- Implementar validaciones automáticas (por ejemplo, en Excel, scripts de revisión o formularios de carga).
- No dejar campos críticos vacíos. En su lugar, usar “No disponible” cuando aplique, y documentarlo en los metadatos.
- Corregir errores ortográficos o tipográficos que generen inconsistencias en valores esperados.

3.6. Criterio de confidencialidad

Definición

La confidencialidad es un principio esencial en la gestión de datos abiertos. Su finalidad es asegurar que la información publicada no vulnere la privacidad de las personas, no comprometa la seguridad institucional ni infrinja la normativa vigente sobre protección de datos personales.

Antes de publicar cualquier conjunto de datos en el portal datos.gov.co, las entidades públicas deben verificar que el contenido no incluya datos personales, sensibles, reservados o clasificados, conforme a lo establecido en la Ley 1581 de 2012 (protección de datos personales), la Ley 1712 de 2014 (Ley de Transparencia y del Derecho de Acceso a la Información Pública) y demás disposiciones aplicables.

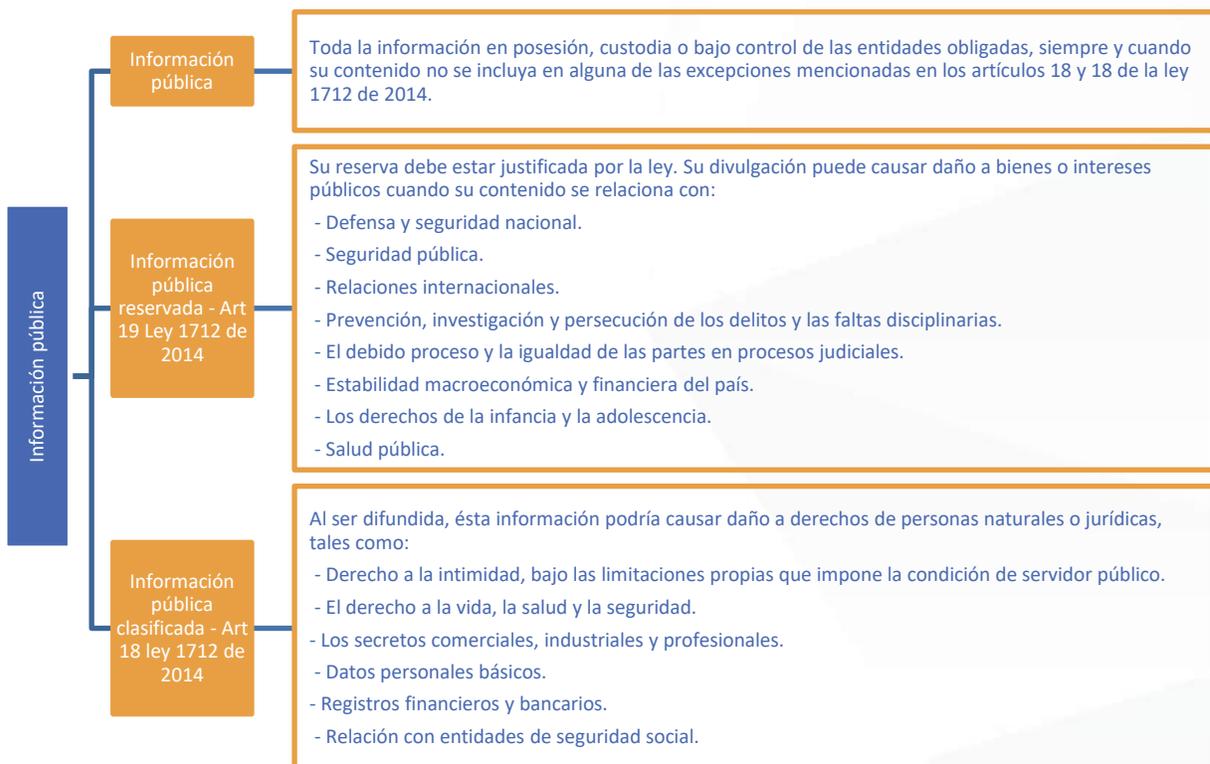
La Ley de Transparencia y del Derecho de Acceso a la Información **ley 1712 de 2014**⁹, la cual fue reglamentada con el **Decreto 103 de 2015**¹⁰, establece de manera explícita que el acceso a la información pública puede ser limitado cuando se trate de Información Pública Clasificada o Reservada. Estas categorías están definidas en el título sobre “Excepciones de acceso a la información” y deben ser cuidadosamente revisadas por las entidades antes de habilitar cualquier conjunto en el portal nacional.

Las siguientes categorías de información restringida deben considerarse al momento de validar el cumplimiento del criterio de confidencialidad:

⁹ <https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=56882>

¹⁰ <https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=60556>

Ilustración 11 Calificación de la información de acuerdo con sus niveles de seguridad



Fuente. Elaboración propia.

⚙️ Aplicación práctica

Antes de publicar un conjunto de datos, las entidades públicas deben:

- **Revisar su contenido** para identificar posibles datos personales o sensibles (nombres, cédulas, direcciones, registros clínicos, etc.).
- **Consultar con el equipo jurídico** si el conjunto está sujeto a alguna restricción normativa.
- **Aplicar anonimización, agregación o supresión** en los campos que puedan comprometer la confidencialidad.
- **Verificar si la información ya ha sido publicada** por otra entidad, evitando duplicidades que puedan generar inferencias indebidas.

Ejemplo aplicado

Caso: El instituto nacional de salud debe publicar el conjunto de datos de casos positivos del covid-19 del país

Situación inicial: El conjunto de datos incluye nombres completos, números de documento, dirección, diagnósticos, barrio y edad.

Acciones recomendadas:

- Eliminar nombres e identificadores personales.
- Reemplazar edades exactas por rangos (por ejemplo: mes y año).
- Clasificar los diagnósticos por categoría (grupo CIE-10).
- Incluir una nota en los metadatos aclarando que el conjunto ha sido anonimizado.

Buenas prácticas institucionales

- Implementar un **protocolo de revisión previa** para identificar riesgos de confidencialidad.
- Documentar el proceso de anonimización o validación legal.
- Utilizar glosarios controlados o niveles de agregación cuando los datos sean sensibles o de baja volumetría.
- Aplicar el **principio de mínima exposición**: publicar solo lo estrictamente necesario para el propósito del conjunto.

3.6.1. Anonimizar los datos

La anonimización es el proceso mediante el cual se transforman o eliminan datos personales y sensibles de un conjunto, para garantizar que ninguna persona o entidad pueda ser identificada directa o indirectamente a partir de la información publicada.

Aunque en el entorno digital actual el anonimato absoluto es difícil de garantizar, la aplicación de buenas prácticas de anonimización **reduce significativamente los riesgos**, protege los derechos fundamentales y aumenta la confianza ciudadana en el ecosistema de datos abiertos.

Una anonimización mal implementada puede exponer información sensible y comprometer la calidad institucional del conjunto, lo que puede derivar en consecuencias legales y reputacionales.

Pasos recomendados para una anonimización efectiva

1. Pre-anonimización

- Evaluar si algún campo o combinación de campos puede llevar a la identificación de personas o entidades.
- Eliminar desde el inicio cualquier dato personal innecesario para el propósito del conjunto.
- Tener especial precaución con **datos sensibles** según la normatividad colombiana.

2. Técnicas de ocultamiento, supresión o seudo-anonimización

Identificar, eliminar o transformar los siguientes tipos de datos antes de su publicación:

- Nombres propios.
- Fechas completas de nacimiento o constitución (excepto año).
- Números de identificación personal (CC, pasaporte, tarjeta de identidad).
- Datos tributarios (NIT, RUT, RUP, RUES).
- Direcciones de domicilio o correo electrónico.
- Números de teléfono, fax o cuentas bancarias.
- Placas de vehículos o identificadores de dispositivos.
- Direcciones IP, números de serie o códigos únicos.
- Cualquier otro identificador personal directo o indirecto.

 **Recomendación:** Cuando sea necesario conservar cierta información estructural, usar agrupaciones, rangos o categorías generales (por ejemplo: rango de edad, municipio sin dirección, mes en lugar de fecha completa).

Referencia técnica recomendada

Para profundizar en las metodologías y técnicas de anonimización de datos estructurados, se recomienda consultar la **Guía para la Anonimización de Datos Estructurados** del DANE (2024). Este documento proporciona orientaciones detalladas sobre el proceso de anonimización, incluyendo: [DANE](#)

- Identificación y clasificación de variables sensibles.
- Evaluación de riesgos de reidentificación.
- Selección y aplicación de técnicas de anonimización.
- Evaluación de la utilidad de los datos anonimizados. La guía está disponible en el siguiente enlace:

 [Guía para la Anonimización de Datos Estructurados - DANE](#)

3.7. Criterio de consistencia

Definición

El criterio de **consistencia** evalúa el grado en que los datos son **coherentes, estandarizados y libres de contradicciones internas** o con respecto a otros conjuntos de datos relacionados. Se refiere tanto a la **homogeneidad en los valores** como a la uniformidad en la **codificación, formato y estructura** de la información.

La consistencia es esencial para garantizar la confiabilidad de los análisis y la interoperabilidad de los datos entre distintas fuentes. Inconsistencias como múltiples formas de escribir un mismo valor o campos duplicados disminuyen la utilidad del conjunto de datos y pueden afectar negativamente su interpretación.

Aplicación práctica

Una entidad pública cumple con este criterio cuando:

- Usa una **codificación uniforme** para los mismos campos (por ejemplo, siempre “CUNDINAMARCA” y no variaciones como “cmarca” o “C/MARCA”).
- Los valores de texto en campos descriptivos tienen **longitud y estructura apropiada**, sin abreviaturas inconsistentes o entradas ambiguas.
- No existen **columnas duplicadas** con diferente nombre pero con contenido similar (por ejemplo, “Municipio” y “Municipio_nombre” con los mismos datos).
- Se utilizan **catálogos o tablas de referencia** para campos repetidos entre conjuntos.

Ejemplo aplicado

Caso: Una entidad publica varios conjuntos de datos sobre cobertura educativa.

Situación problemática:

- En uno aparece “Bogotá D.C.”, en otro “Bogota”, y en otro “BOG” como nombre de la ciudad.
- El campo “NivelEducativo” aparece en dos columnas distintas con los mismos valores.

Acciones recomendadas:

- Estandarizar los valores a una sola nomenclatura validada, idealmente usando códigos DANE u otro estándar nacional.
- Consolidar columnas duplicadas y corregir diferencias en mayúsculas, tildes o abreviaciones.
- Aplicar validaciones en Excel, bases de datos o scripts para detectar variaciones del mismo valor.

La siguiente imagen muestra inconsistencias en los valores de una columna, evidenciando múltiples formas de registrar el mismo concepto (“Código Producto”), lo cual impacta la calidad del conjunto.

Ilustración 12 Conjunto de datos con datos inconsistentes

ID	Departamento	Fecha Registro	Código Producto
1	CUNDINAMARCA	2023-01-15	P001
2	cmarca	15/01/2023	P001
3	C/MARCA	2023-01-15	P1
4	CUNDINAMARCA	15-Jan-2023	P001
5	Cundinamarca	2023/01/15	001P
6	cundinamrca	2023-01-15	P001
7	CUNDINAMARCA	15-01-2023	001-P
8	c/Marca	2023-01-15	P001

Fuente. Captura de pantalla de datos.gov.co

✓ Buenas prácticas institucionales

- Crear tablas de referencia internas para campos como municipio, departamento, tipo de trámite, modalidad de contrato, etc.
- Aplicar reglas de validación antes de publicar, detectando diferencias en formato, escritura o longitud de los valores.
- Unificar columnas que contengan información duplicada o redundante.
- Promover la revisión entre equipos de diferentes áreas para detectar posibles incoherencias en conjuntos relacionados.

3.8. Criterio de credibilidad

✚ Definición

La **credibilidad** es el grado en que los datos se consideran **confiables, ciertos y verificables** dentro de un contexto específico de uso. Según la norma **ISO/IEC 2501214¹¹**, un conjunto de datos es creíble cuando incluye elementos que respaldan

¹¹ <https://iso25000.com/index.php/normas-iso-25000/iso-25012>

su origen, autoría y validez, tales como: documentación, normatividad aplicable, contacto institucional y metadatos completos.

Este criterio permite al usuario final **confiar en la fuente y calidad del conjunto de datos**, y facilita la trazabilidad institucional sobre su creación, mantenimiento y publicación.

⚙️ Aplicación práctica

Una entidad pública cumple con este criterio cuando:

- Los metadatos incluyen información sobre la **fuentes de los datos**, su base normativa, objetivo y cobertura.
- Se identifica claramente la **entidad publicadora** y se ofrece un **correo electrónico institucional** de contacto.
- Se diligencian adecuadamente los **campos del editor** en el portal datos.gov.co, reflejando el carácter oficial de la publicación.
- Las **columnas del conjunto** cuentan con descripciones válidas, detalladas y sin errores.

📌 Importante: La ausencia de un contacto institucional válido o la presencia de campos vacíos o genéricos en los metadatos puede hacer que un conjunto de datos creíble en su origen no sea percibido como tal por los usuarios.

💡 Ejemplo aplicado

Caso: Una entidad publica un conjunto sobre inversión en infraestructura pública.

Situación recomendada:

- La descripción incluye el programa del cual provienen los datos, su respaldo normativo y el año de referencia.
- En los metadatos figura la entidad publicadora y un correo institucional como contacto (ej. datos@entidad.gov.co).
- Las columnas “MontoInvertido”, “FechaInicio” y “ProyectoID” tienen descripciones comprensibles y completas.

La siguiente imagen muestra algunos de los campos clave en los metadatos del portal datos.gov.co que se deben diligenciar para garantizar la trazabilidad y la confianza institucional del conjunto de datos.

Ilustración 13 Metadatos Credibilidad

Editar los metadatos ×

Licencia de Atribución

Tipo de licencia (si procede)
-- Sin licencia --

Datos suministrados por
MinTIC

Enlace de la fuente
<https://www.datos.gov.co/>

Email de contacto

Dirección electrónica
datosabiertos@mintic.gov.co

Nombre del recurso

Nombre del recurso
nombre del ejemplo

Fuente. Captura de pantalla de [datos.gov.co](https://www.datos.gov.co)

✓ Buenas prácticas institucionales

- Utilizar correos institucionales genéricos como punto de contacto, que puedan mantenerse en el tiempo y ser atendidos por equipos, no personas individuales.
- Verificar que el nombre del editor institucional esté completo y represente oficialmente a la entidad.
- Incluir siempre la fuente, normativa o contexto legal del conjunto, así como los métodos de recolección o procesamiento de los datos.
- Configurar perfiles de usuario de carga con nombres institucionales e imágenes corporativas visibles.

3.9. Criterio de disponibilidad

✚ Definición

La **disponibilidad** es el grado en que un conjunto de datos está **accesible, actualizado y operativo para ser consultado por los usuarios**. Este criterio combina dos atributos

fundamentales: la **accesibilidad** (facilidad de acceso y consulta) y la **actualidad** (vigencia de la información), reconociendo que los datos solo son útiles si pueden encontrarse y están actualizados.

Este enfoque reconoce que la simple existencia de datos en línea no es suficiente: deben estar bien organizados, técnicamente accesibles y reflejar información vigente.

⚙️ **Aplicación práctica**

Una entidad pública cumple con este criterio cuando:

- Publica sus datos en un **portal de acceso público** como datos.gov.co, sin requerir registro o permisos especiales.
- Los **metadatos están diligenciados** correctamente, incluyendo título, descripción, frecuencia de actualización, fuente y contacto institucional.
- El conjunto cuenta con **etiquetas de búsqueda**, enlaces a documentación o normativa, y está disponible en formatos abiertos.
- La **fecha de última actualización** es coherente con la **frecuencia declarada**.
- La plataforma o sistema donde se alojan los datos funciona correctamente y sin caídas frecuentes.

💡 **Ejemplo aplicado**

Caso: Una entidad territorial publica datos sobre contrataciones públicas mensuales.

Situación óptima:

- El conjunto aparece en el portal nacional sin necesidad de registro.
- Los metadatos incluyen etiquetas, descripción, enlace a la normativa y correo de contacto.
- La fecha de actualización es reciente y está alineada con la frecuencia mensual declarada.

✅ **Buenas prácticas institucionales**

- Validar que todos los conjuntos estén **operativos y descargables** desde el portal.
- Establecer mecanismos para **monitorear el cumplimiento de la frecuencia de actualización**.
- Garantizar que los datos estén disponibles incluso fuera de horarios administrativos (alta disponibilidad 24/7).

- Revisar periódicamente enlaces, formatos y funcionamiento general del conjunto desde la perspectiva de un usuario externo.

3.10. Criterio de eficiencia

Definición

La **eficiencia** se refiere a la capacidad del conjunto de datos para ser **procesado, descargado y analizado de manera rápida y sin redundancias innecesarias**. Este criterio se relaciona con el diseño interno del conjunto, incluyendo su estructura, la ausencia de duplicados, y su completitud, ya que estos factores impactan directamente el rendimiento de la plataforma y la experiencia del usuario.

En versiones anteriores de la guía, la eficiencia se consideraba cumplida por defecto. Sin embargo, en la presente versión se evalúa como una **característica asociada a la calidad estructural del conjunto**, considerando la redundancia de registros y columnas, así como la completitud de la información.

Aplicación práctica

Una entidad pública cumple con este criterio cuando:

- El conjunto **no contiene columnas duplicadas o redundantes**, es decir, con el mismo contenido bajo distinto nombre.
- **No hay filas duplicadas**, especialmente en campos clave como identificadores, fechas, o nombres.
- La **estructura del conjunto está optimizada** para facilitar la descarga, visualización y análisis sin sobrecargar recursos computacionales.
- Se evita publicar grandes cantidades de datos repetitivos o irrelevantes para el propósito del conjunto.

Ejemplo aplicado

Caso: Una entidad publica datos sobre compras institucionales.

Situación problemática:

- El conjunto incluye dos columnas: “Producto” y “Producto 1”, con los mismos valores.

- Se observan múltiples registros idénticos con el mismo “Número de Compra” y “Fecha”.

Acciones recomendadas:

- Consolidar las columnas duplicadas en una sola.
- Eliminar filas idénticas o validar si se trata de errores de carga.
- Verificar que cada fila represente una observación única y relevante.

La siguiente imagen muestra un conjunto de datos con columnas duplicadas y filas repetidas, lo que reduce su eficiencia y puede generar sobrecarga o errores al momento de análisis automatizado.

Ilustración 14 Error en eficiencia

ID	Nombre	Producto	Cantidad	Precio Unitario	Total	Fecha de Compra	Producto 1
1	Juan	P1	10	100	1,000	2023 Jan 01 12:00:00 AM	P1
2	María	P2	5	200	1,000	2023 Jan 02 12:00:00 AM	P2
3	Carlos	P3	8	150	1,200	2023 Jan 03 12:00:00 AM	P3
4	Ana	P4	2	300	600	2023 Jan 04 12:00:00 AM	P4
5	Luis	P5	7	250	1,750	2023 Jan 05 12:00:00 AM	P5
6	Sofía	P6	4	400	1,600	2023 Jan 06 12:00:00 AM	P6
7	Elena	P7	6	350	2,100	2023 Jan 07 12:00:00 AM	P7

Fuente. Captura de pantalla de datos.gov.co

✓ Buenas prácticas institucionales

- Revisar los conjuntos de datos con herramientas de validación para detectar duplicados (Excel, scripts en Python, herramientas del portal).
- Mantener una estructura clara, con columnas relevantes, bien tituladas y sin repeticiones.
- Aplicar un control de calidad antes de publicar, identificando redundancias o registros innecesarios.
- Si el conjunto contiene múltiples hojas u hojas ocultas, consolidar todo en un solo archivo o eliminar contenido residual.

3.11. Criterio de exactitud

Definición

La **exactitud** es el grado en que los datos **representan fielmente el valor real** del atributo o concepto que pretenden describir. Según la norma **ISO/IEC 25012:2008**, implica que los datos no tengan errores, distorsiones o deformaciones respecto a la realidad que buscan modelar.

En el contexto de los datos abiertos en Colombia, este criterio abarca dos dimensiones complementarias:

- **Exactitud sintáctica:** Evalúa si los datos cumplen con reglas formales de escritura, ortografía, formato y estructura (por ejemplo, uso correcto de tildes, mayúsculas, puntuación, unidades de medida, etc.).
- **Exactitud semántica:** Evalúa si los valores representan adecuadamente su significado dentro del contexto previsto. Por ejemplo, que “municipio” se refiera efectivamente a una entidad territorial válida y no a un nombre ficticio o confuso.

Aplicación práctica

Una entidad pública cumple con este criterio cuando:

- Los valores de texto están **bien escritos, sin errores ortográficos ni abreviaciones arbitrarias**.
- Se respeta la **gramática y formato** establecido para nombres, lugares, fechas y conceptos clave.
- Se han normalizado los datos de entrada, eliminando diferencias innecesarias en mayúsculas, tildes o puntuación.
- Se evita incluir datos mal escritos, fragmentados, en otros idiomas o con estructuras incorrectas.
- Se han aplicado **procesos de revisión léxica o semántica** que garanticen la coherencia del contenido.

Ejemplo aplicado

Caso: Una entidad publica un conjunto sobre sanciones disciplinarias.

Situación problemática:

- En la columna “Entidad sancionadora” se encuentran valores como “Procuraduría”, “procuraduria”, “PROCURADURIA” y “Proc.” como si fueran equivalentes.
- En otra columna aparece “Ministerio d Salud” (con errores de ortografía).

Acciones recomendadas:

- Unificar todos los valores a “Procuraduría General de la Nación”.
- Aplicar corrección ortográfica automática en nombres institucionales.
- Revisar que todas las entradas estén en idioma español, con estructura y longitud adecuada.

La siguiente imagen muestra ejemplos de campos con errores ortográficos, valores redundantes y abreviaciones no estandarizadas, lo que impacta negativamente el nivel de exactitud del conjunto.

Ilustración 15 Revisión de metadatos para el criterio de exactitud sintáctica

Vista previa de la tabla Ver datos Crear visualización

NOMBRE PRESTADOR	DIRECCION	BARRIO	CLASE PRESTADOR	LOCALIZACION
Carmen teresa cárdenas Rodri...	CARRERA 3 # 13-12 LOCAL 56...	EL TREBOL	Profesional Independiente	(4.711437°, -74.223211°)
Carola Andrea Ortegón Díaz	kr. 3 a no. 13 -12 centro comer...	el trebol	profesional independiente	(4.71115°, -74.22295°)
Carola Andrea Ortegón Díaz	diagonal 19 no. 3 c -05	villa nueva	profesional independiente	(4.7131°, -74.22055°)
catalina Alejandra Rodríguez ...	calle 5 # 02-73 cs 302	centro	profesional independiente	(4.70697°, -74.2292°)
catalina Alejandra Rodríguez ...	calle10 n. 16b -15	poblado	Profesional Independiente	(4.7217°, -74.2292°)
centro de Intervención idear s...	Cra 2 N° 3 - 16 oficina 101	Centro	Profesional Independiente	(4.70538°, -74.22952°)

Fuente. Captura de pantalla de datos.gov.co

✓ Buenas prácticas institucionales

- Utilizar correctores ortográficos automatizados antes de publicar conjuntos.
- Validar campos clave con herramientas de procesamiento de lenguaje natural (PLN) adaptadas al español.
- Estandarizar nombres institucionales, territoriales y temáticos, usando diccionarios o catálogos oficiales.
- Aplicar reglas de formato para fechas, códigos, números y siglas.
- Revisar columnas de texto con alta variabilidad para identificar entradas redundantes o incorrectas.

3.12. Criterio de portabilidad

La **portabilidad** evalúa la facilidad con la que un conjunto de datos puede ser **descargado, transferido, procesado o reutilizado en distintos entornos tecnológicos**, sin pérdida de estructura ni integridad. De acuerdo con la norma **ISO/IEC 25010:2008**, se trata del grado en que los datos pueden ser trasladados entre sistemas conservando su funcionalidad.

Este criterio está directamente relacionado con la **interoperabilidad tecnológica**, permitiendo que los datos puedan ser utilizados por múltiples herramientas, lenguajes de programación o plataformas de análisis.

Aplicación práctica

Una entidad pública cumple con este criterio cuando:

- Publica los datos en **formatos abiertos y estructurados**, tales como CSV, JSON, XML, TSV, RDF, entre otros, compatibles con distintos sistemas.
- Evita el uso de **caracteres especiales** en los nombres de archivo o en campos clave que puedan causar errores de lectura o ejecución.
- Minimiza la presencia de **valores nulos**, lo cual favorece una experiencia más estable de análisis y reutilización.
- **Promueve el uso de la API del portal de datos abiertos** para acceder a conjuntos de gran volumen, facilitando su consulta, filtrado y procesamiento sin necesidad de descargar archivos pesados.
- Documenta en los metadatos cómo acceder al conjunto vía API y qué formatos están disponibles.

 **Importante:** La portabilidad **no implica reducir el tamaño ni segmentar conjuntos de datos grandes y completos**. En el caso de datos nacionales o de alto volumen, se recomienda el uso de la **API Socrata** del portal datos.gov.co, que permite la integración eficiente de estos conjuntos en entornos analíticos.

Ejemplo aplicado

Caso: Una entidad publica el registro nacional de contratos.

Prácticas recomendadas:

- El conjunto está disponible en formato CSV y JSON.

- Los nombres de columnas y archivos no incluyen caracteres especiales.
- Se especifica en los metadatos la posibilidad de consultar el conjunto a través de API.
- No se fragmenta innecesariamente el conjunto, asegurando su completitud.

Ilustración 16 Formatos para la publicación de datos abiertos



Fuente. Elaboración propia

✓ Buenas prácticas institucionales

- Publicar conjuntos en **formatos abiertos y estandarizados** soportados por el portal.
- Documentar claramente el uso de la **API** cuando el volumen de datos lo amerite.
- Evitar errores técnicos derivados de caracteres especiales o estructuras no compatibles.
- **No segmentar artificialmente** los conjuntos por volumen. En su lugar, facilitar el acceso programático mediante endpoints bien definidos.

3.13. Criterio de precisión

Definición

La **precisión** evalúa qué tan bien representan los datos las **entidades reales del mundo**, sin ambigüedades ni simplificaciones excesivas. Según la norma **ISO/IEC 25012:2008**, la precisión se refiere al grado en que los datos coinciden con el valor real que deberían reflejar, en un contexto específico de uso.

Este criterio es clave para asegurar que los datos abiertos sean confiables, útiles y relevantes para análisis, investigaciones o sistemas basados en datos (como inteligencia artificial, modelado predictivo o visualizaciones complejas).

Aplicación práctica

Una entidad pública cumple con este criterio cuando:

- Las **columnas del conjunto de datos** tienen suficiente **diversidad y variabilidad** para representar fielmente la realidad (por ejemplo: no repetir el mismo valor en toda la columna).
- Se evitan columnas “vacías en contenido”, es decir, aquellas que contienen un solo valor repetido sin ofrecer diferenciación útil.
- Se identifican y eliminan **columnas poco informativas o sin valor analítico**, como aquellas con valores constantes.
- Se aseguran valores **numéricos o categóricos suficientemente detallados** que representen rangos, categorías, zonas geográficas, fechas o indicadores con precisión adecuada.

Ejemplo aplicado

Caso: Una entidad publica un conjunto de datos sobre coberturas de salud por región.

Situación problemática:

- En la columna “Departamento” aparece el mismo valor en todas las filas: “CUNDINAMARCA”.
- En la columna “PorcentajeCobertura” todos los valores son “100”.

Acciones recomendadas:

- Revisar si la información está mal segmentada (por ejemplo, si falta incluir otros departamentos).
- Evaluar si la columna “PorcentajeCobertura” refleja una medición real o es un valor de carga por defecto.
- Eliminar columnas que no aporten variabilidad y, por tanto, precisión.

La siguiente imagen muestra un conjunto de datos donde múltiples columnas presentan baja diversidad de valores, afectando la calidad de análisis y la representatividad del contenido.

Ilustración 17 Ejemplo error precisión

ID	Edad (años)	Ingreso Mensual (\$)	Ciudad de Residencia	Estado Civil
1	25	3,000	Bogotá	Soltero
2	25	3,000	Bogotá	Soltero
3	40	4,500	Medellín	Casado
4	35	3,200	Cali	Soltero
5	35	3,200	Cali	Soltero
6	29	2,800	Bogotá	Casado
7	29	2,800	Bogotá	Casado
8	40	4,700	Medellín	Casado

Fuente. Captura de pantalla de datos.gov.co

✓ Buenas prácticas institucionales

- Verificar que cada columna tenga **valores distintos y suficientes** según su naturaleza (fechas, regiones, porcentajes, etc.).
- Evaluar si los campos numéricos presentan **variabilidad real** y no valores fijos.
- No incluir columnas que contengan **un único valor repetido**, ya que no aportan a los análisis ni a la utilidad del conjunto.
- Documentar en los metadatos si ciertos campos tienen valores homogéneos debido al diseño del conjunto (por ejemplo, subconjuntos por región o período).

3.14. Criterio de recuperabilidad

Definición

La **recuperabilidad** mide la **capacidad de un conjunto de datos para ser encontrado, comprendido y reutilizado en el tiempo**, gracias a una adecuada gestión de sus metadatos y al seguimiento de su trazabilidad y actualización.

Este criterio considera no solo el acceso actual al conjunto de datos, sino también la posibilidad de **auditar su historial, entender su origen, y preservar su funcionalidad operativa** en plataformas, sistemas o aplicaciones futuras. En esencia, la recuperabilidad garantiza que el conjunto **siga siendo útil, interpretable y accesible con el paso del tiempo**.

Aplicación práctica

Una entidad pública cumple con este criterio cuando:

- Sus conjuntos de datos cuentan con **metadatos completos y bien estructurados**, incluyendo título, descripción, etiquetas, contacto institucional, frecuencia de actualización, normativa asociada y descripción de cada campo.
- Los metadatos permiten **auditar el acceso y actualización** del conjunto, facilitando la trazabilidad institucional (por ejemplo, saber cuándo y por quién fue modificado).
- La información está disponible en formatos accesibles y a través de portales que permitan la descarga y consumo mediante API.
- Se ha documentado claramente el ciclo de vida del conjunto, su propósito, su origen y los criterios bajo los cuales se actualiza.

Ejemplo aplicado

Caso: Una entidad publica datos sobre programas sociales activos.

Situación recomendada:

- En los metadatos se identifica el responsable de publicación, la fecha de última actualización, la normativa que respalda el programa, y un correo de contacto.
- El conjunto incluye descripciones detalladas de cada campo, etiquetas claras y términos normalizados.
- Se puede verificar cuándo se modificó por última vez el conjunto y con qué periodicidad se actualiza.

✓ Buenas prácticas institucionales

- Diligenciar todos los campos de metadatos disponibles en el portal de datos abiertos.
- Establecer procedimientos para el control de versiones y auditoría del conjunto.
- Mantener la coherencia entre los metadatos y el contenido del archivo publicado.
- Publicar la descripción de los campos directamente en los metadatos, no solo en archivos separados.
- Asegurar que los datos se puedan consultar tanto por descarga directa como a través de APIs.

3.15. Criterio de relevancia

✚ Definición

La **relevancia** mide el grado en que los datos publicados por una entidad **responden a las necesidades de información de su sector, del público y de otros actores institucionales**. Este criterio refleja si los datos aportan valor para la toma de decisiones, la formulación de políticas, el control social o el análisis sectorial, de acuerdo con las prioridades temáticas definidas a nivel nacional e internacional.

🌐 Referentes

Este criterio está alineado con:

- La **Hoja de Ruta Nacional y/o sectoriales de Datos Abiertos estratégicos y/o el plan nacional de desarrollo vigente**, que establecen criterios de priorización por sector y propósito institucional.
- Índices internacionales de datos abiertos como el **Open Data Barometer**, el **Global Open Data Index** y el **Our Data Index**, que identifican conjuntos prioritarios como: presupuesto, contratación pública, resultados electorales, educación, salud, transporte, estadísticas ambientales, comercio, propiedad de la tierra, entre otros.
- **Conpes o proyectos sectoriales de tecnologías emergentes** que establecen la necesidad de disponer datos

⚙️ Aplicación práctica

Una entidad pública cumple con este criterio cuando:

- El conjunto de datos incluye una **categoría temática válida y alineada con su sector** (como Cultura, Educación, Justicia, Salud, etc.).
- La **descripción del conjunto es coherente** con la categoría asignada y hace referencia explícita a temas de interés público.
- El conjunto tiene **un volumen mínimo de registros** (por ejemplo, más de 50 filas), lo que permite realizar análisis comparativos o temporales.
- Se priorizan temas con **alta demanda pública o valor estratégico** para la entidad.

✦ Publicar un conjunto con una sola fila o con descripciones ambiguas reduce su relevancia institucional y limita su uso en procesos de análisis.

💡 Ejemplo aplicado

Caso: Una entidad ambiental publica datos sobre calidad del aire.

Situación ideal:

- El conjunto tiene la categoría “Medio ambiente”.
- La descripción incluye palabras clave como “PM10”, “monitoreo”, “ciudades capitales”, “normativa ICA”.
- El archivo contiene más de 5.000 registros históricos, con al menos 10 columnas de variables relevantes.

Situación deficiente:

- La categoría no está asignada o aparece como “Otros”.
- La descripción solo dice “Base de datos de calidad”.
- El conjunto tiene una sola fila o valores nulos en la mayoría de campos.

La siguiente imagen muestra un conjunto con información mínima, lo que impide su análisis y lo hace poco relevante.

Ilustración 18 Muestra de conjunto de datos con información no relevante

ID	Nombre	Producto	Cantidad	Precio Unitario	Total	Fecha de Compra	Producto 1
1	Juan	P1	10	100	1,000	2023 Jan 01 12:00:00 AM	P1

Fuente. Captura de pantalla de datos.gov.co

✓ Buenas prácticas institucionales

- Asignar siempre una **categoría temática válida** al momento de la publicación.
- Redactar descripciones completas que expliquen el propósito, cobertura y utilidad del conjunto.
- Priorizar la publicación de datos de interés público, conforme a las agendas sectoriales y los compromisos institucionales.
- Garantizar que el conjunto tenga un volumen mínimo y columnas suficientes para ser analizado

3.16. Criterio de trazabilidad

✚ Definición

La **trazabilidad** mide la capacidad de un conjunto de datos para ser **rastreado, auditado y actualizado en el tiempo**, garantizando la transparencia, la confianza institucional y la posibilidad de reutilización. De acuerdo con la norma **ISO 9001:2008**, este atributo permite identificar los cambios aplicados al conjunto, su origen, sus responsables y las fechas clave de modificación.

En la práctica, la trazabilidad requiere que los datos sean actualizables, que cuenten con metadatos completos desde su publicación y que se documenten las modificaciones a lo largo del tiempo.

Este criterio es especialmente importante para la **interoperabilidad entre entidades** y para la **gestión de calidad en los procesos de apertura y mantenimiento de datos**.

Aplicación práctica

Una entidad pública cumple con este criterio cuando:

- Todos los campos de **metadatos requeridos están diligenciados**, incluyendo título, descripción, categoría, cobertura, frecuencia, normativa aplicable, contacto institucional y responsable de publicación.
- Se especifica la **fecha de última actualización** y esta es coherente con la frecuencia de publicación indicada.
- Se incluye un **correo electrónico institucional** de contacto o responsable, permitiendo la trazabilidad del proceso.
- **No se incluyen fechas explícitas en el título del conjunto**, evitando que este quede obsoleto o se perciba como no actualizable.

Ejemplo aplicado

Caso correcto:

Un conjunto de datos titulado “Inventario Nacional de Humedales” incluye:

- Título neutral sin fecha.
- Frecuencia de actualización anual.
- Fecha de última actualización: enero de 2024.
- Metadatos completos.
- Contacto: biodiversidad@minambiente.gov.co.

Caso incorrecto:

- Título: “Inventario de Humedales 2021”.
- No contiene responsable de publicación ni correo institucional.
- Última actualización desconocida.

Buenas prácticas institucionales

- Completar todos los metadatos desde la publicación inicial.

- Evitar títulos con fechas (estas deben ir en los campos de frecuencia o cobertura temporal).
- Incluir contacto institucional (no personal), idealmente un buzón funcional.
- Documentar cada actualización del conjunto y reflejarlo en la fecha correspondiente.

Las siguientes imágenes pueden insertarse para reforzar visualmente los errores comunes y las acciones correctas.

Ilustración 13 Muestra de conjunto de datos con error en el título

Tipo	Nombre	Acciones	Último actualizado	Categoría	Propietario	Destinatarios
	Errores Año 2024 Errores	...	December 21, 2024	Ciencia, Tecnología e Innovación	 Ministerio de Tecnologías de la Información y las Comunicaciones-MINTIC	 Público

Fuente. Captura de pantalla de datos.gov.co

Ilustración 14 Metadatos a diligenciar



Fuente. Captura de pantalla de datos.gov.co

3.17. Criterio de unicidad

Definición

La **unicidad** mide el grado en que los datos de un conjunto **no contienen registros o campos duplicados**, tanto a nivel de **filas** como de **columnas clave**. Este criterio garantiza la **integridad estructural de los datos**, evitando redundancias que pueden distorsionar resultados, afectar análisis estadísticos y restar confiabilidad al conjunto.

Aplicación práctica

Una entidad pública cumple con este criterio cuando:

- Cada fila del conjunto representa una **observación única** y no existen registros repetidos (por ejemplo, dos veces el mismo número de identificación o la misma fecha de transacción).
- No hay **columnas duplicadas** con nombres similares o contenidos redundantes (por ejemplo, “FechaCompra” y “Fecha de Compra 1” con los mismos valores).
- Se han aplicado controles de calidad antes de publicar, incluyendo revisión de duplicados, estandarización de campos y verificación de registros clave.
- La estructura del conjunto evita confusiones, errores de carga o ambigüedad en la interpretación.

Ejemplo aplicado

Caso: Una entidad publica datos sobre licencias de construcción.

Situación ideal:

- Cada fila corresponde a una única licencia, con identificador no repetido.
- Las columnas tienen nombres únicos y representaciones claras (sin duplicación semántica).
- Se ha validado previamente la eliminación de duplicados mediante scripts o funciones en Excel.

Situación deficiente:

- Existen varias filas con el mismo número de licencia, fecha y dirección.
- Se repite la columna “Área construida” con nombres ligeramente distintos.
- Esto puede generar doble conteo o errores de análisis.

Ilustración 19 Evidencia de filas y columnas duplicadas

ID	Nombre	Producto	Fecha de Compra	Producto 1	Fecha de Compra 1
1	Juan	P1	2023 Jan 01 12:00:00 AM	P1	2023 Jan 01 12:00:00 AM
2	María	P2	2023 Jan 02 12:00:00 AM	P2	2023 Jan 02 12:00:00 AM
3	Carlos	P3	2023 Jan 03 12:00:00 AM	P3	2023 Jan 03 12:00:00 AM
4	Ana	P4	2023 Jan 04 12:00:00 AM	P4	2023 Jan 04 12:00:00 AM
5	Juan	P1	2023 Jan 01 12:00:00 AM	P1	2023 Jan 01 12:00:00 AM
1	Juan	P1	2023 Jan 01 12:00:00 AM	P1	2023 Jan 01 12:00:00 AM
6	Sofía	P5	2023 Jan 05 12:00:00 AM	P5	2023 Jan 05 12:00:00 AM
2	María	P2	2023 Jan 02 12:00:00 AM	P2	2023 Jan 02 12:00:00 AM
7	Luis	P6	2023 Jan 06 12:00:00 AM	P6	2023 Jan 06 12:00:00 AM

Fuente. Captura de pantalla de datos.gov.co

✓ Buenas prácticas institucionales

- Aplicar filtros o validaciones para detectar duplicados antes de publicar el conjunto.
- Establecer campos clave (por ejemplo, número de contrato, NIT, código DANE) como identificadores únicos.
- Eliminar columnas que sean equivalentes o resultado de errores de exportación (“variable1”, “variable1_1”).
- Documentar en los metadatos si el conjunto tiene estructura longitudinal (por ejemplo, si una persona aparece más de una vez porque hay varios eventos).

4. Lineamientos para preparar conjuntos de datos habilitantes para IA

4.1. ¿Qué es un conjunto de datos habilitante para IA?

Un conjunto de datos habilitante para inteligencia artificial (IA) es aquel que posee características técnicas, semánticas y legales que lo hacen adecuado para ser utilizado como insumo en el entrenamiento, validación, despliegue y supervisión de sistemas de IA, especialmente aquellos basados en modelos de aprendizaje automático (machine learning) y procesamiento de lenguaje natural (NLP).

A diferencia de un conjunto de datos convencional, un dataset habilitante para IA no solo debe estar disponible en formatos abiertos y con licencia reutilizable, sino que debe cumplir con requisitos avanzados de calidad, granularidad, completitud y trazabilidad que permitan su procesamiento automático sin intervención intensiva de limpieza o transformación manual. Estos conjuntos de datos deben estar preparados para ser leídos por máquinas, pero también bien documentados para facilitar su comprensión por parte de los desarrolladores, analistas de datos, y auditores de algoritmos.

Características clave de un conjunto de datos habilitante para IA:

- **Estructura normalizada y consistente:** Los datos deben estar organizados de forma estructurada, sin ambigüedad en los campos ni duplicidad en las columnas. Deben evitarse celdas combinadas o estructuras semicomplejas que dificulten su lectura automática.
- **Metadatos completos y actualizados:** Es indispensable contar con metadatos que describan con claridad el contenido, origen, periodicidad, actualizaciones, responsable del conjunto de datos y condiciones de uso. Estos metadatos son esenciales para garantizar trazabilidad y confiabilidad, así como para permitir la interoperabilidad entre sistemas.
- **Cobertura suficiente y granularidad adecuada:** Los datos deben cubrir una muestra representativa del fenómeno que se desea modelar, con un volumen de registros que permita el análisis estadístico y la detección de patrones. Debe evitarse la

segmentación excesiva o la publicación de subconjuntos mínimos que limiten la aplicabilidad del dataset.

- **Calidad intrínseca y calidad contextual:** Conforme a la norma ISO/IEC 25012, los datos deben ser precisos, completos, consistentes, actualizados y relevantes para el contexto de aplicación de la IA. Además, deben ser accesibles bajo condiciones legales y técnicas adecuadas para su tratamiento automatizado.
- **Ausencia de sesgos evidentes o errores sistemáticos:** Para evitar que los sistemas de IA repliquen o amplifiquen inequidades, es necesario que los datos no estén marcadamente sesgados, incompletos o incorrectamente clasificados. Esto implica una revisión previa de los criterios de inclusión, la representatividad y la integridad del conjunto.
- **Licenciamiento abierto y claro:** El conjunto debe contar con licencias que permitan explícitamente su uso, adaptación y redistribución para fines de desarrollo tecnológico, incluyendo inteligencia artificial. En Colombia, la licencia estándar aplicada es la Creative Commons Atribución – Compartir Igual 4.0 Internacional (CC BY-SA 4.0), la cual cumple con estos requisitos y garantiza compatibilidad internacional.

4.2. ¿Cómo identificar o construir un conjunto de datos habilitante para IA en una entidad pública?

Para que una entidad pública pueda contribuir de manera estratégica al ecosistema nacional de inteligencia artificial, es fundamental que identifique y fortalezca aquellos conjuntos de datos que puedan servir como insumo habilitante para el desarrollo de soluciones algorítmicas. Esta tarea no requiere iniciar desde cero, sino más bien reconocer, adaptar y mejorar los activos de datos que ya existen dentro de la entidad.

Etapas para identificar o construir conjuntos habilitantes:

1. Diagnóstico de activos de datos existentes

El primer paso consiste en realizar un inventario de los conjuntos de datos disponibles en la entidad, priorizando aquellos que cumplen con las siguientes características:

- Contienen registros estructurados y actualizables.
- Representan fenómenos críticos o procesos misionales.
- Están vinculados a funciones con alto potencial de automatización, predicción o análisis avanzado.
- Esta revisión puede apoyarse en el catálogo interno de datos, el registro de bases para interoperabilidad, y la hoja de ruta sectorial de datos abiertos.

2. Evaluación de criterios de calidad e interoperabilidad

Una vez identificados los conjuntos con potencial, se debe aplicar la evaluación de calidad descrita en esta guía para identificar brechas en:

- Completitud y validez de los datos.
- Consistencia estructural y semántica.
- Nivel de documentación de los metadatos.
- Existencia de duplicados, valores nulos o errores sistemáticos.
- La herramienta de evaluación automatizada del portal datos.gov.co permite obtener un diagnóstico técnico objetivo con base en los 16 criterios de calidad establecidos.

3. Ajustes técnicos y fortalecimiento del conjunto

Los conjuntos seleccionados deben ser ajustados para mejorar su aptitud para IA. Esto puede incluir:

- Limpieza de duplicados y estandarización de campos.
- Inclusión de variables clave necesarias para el análisis (fechas, ubicaciones, códigos normalizados).
- Ampliación del volumen de registros o frecuencia de actualización.
- Mejora de los metadatos, incluyendo definiciones de campos, origen,

licenciamiento y responsables.

En esta etapa se recomienda la colaboración entre equipos de TI, estadística, jurídico y áreas misionales, para asegurar una publicación robusta, ética y legalmente sólida.

4. Evaluación del potencial de uso en IA

Para validar que el conjunto puede ser efectivamente usado como insumo para IA, se debe verificar que:

- La información sea representativa del fenómeno que se desea modelar.
- La granularidad permita entrenar algoritmos o generar inferencias.
- El conjunto no esté sesgado, incompleto o limitado por restricciones normativas.
- Exista una demanda o caso de uso concreto identificado por la entidad o por el ecosistema de innovación.

5. Publicación con enfoque habilitante

La publicación del conjunto debe hacerse en el portal de datos abiertos con:

- Metadatos completos.
- Documentación técnica accesible.
- Licencia abierta.
- Formato estructurado y reutilizable (por ejemplo: CSV, JSON, interoperabilidad vía API).
- Acompañamiento con visualizaciones, ejemplos de uso o sintaxis de consulta, si es posible.

4.3. Recomendaciones para la publicación de conjuntos de datos abiertos con potencial para IA

Para que los conjuntos de datos abiertos sean aprovechados efectivamente en desarrollos de inteligencia artificial, es necesario que las entidades adopten prácticas

adicionales durante su publicación. Estas recomendaciones no son obligatorias, pero sí altamente sugeridas para mejorar la reutilización avanzada de los datos:

4.3.1. Documentación y metadatos enriquecidos

- Además de los metadatos requeridos en el portal, se recomienda incluir información complementaria como:
 - Diccionario de datos detallado.
 - Esquema de relaciones entre tablas si existen múltiples archivos.
 - Frecuencia de actualización prevista y versión de los datos.
 - Fuentes originales de los datos y metodología de recolección.
- Incluir una sección con supuestos, limitaciones y contexto de los datos es clave para evitar interpretaciones erróneas en modelos automatizados.

4.3.2. Formatos estructurados y consistentes

- Priorizar formatos abiertos y estructurados como CSV, JSON o Parquet, que pueden ser procesados fácilmente por algoritmos de aprendizaje automático.
- Evitar formatos cerrados, comprimidos sin estándar o presentaciones en PDF u hojas escaneadas.

4.3.3. Identificación de campos clave

- Señalar explícitamente las variables objetivo, categóricas, numéricas y sensibles, cuando aplique.
- Indicar campos con identificadores únicos y aquellos con funciones de referencia cruzada o geoespacial.

4.3.4. Reducción de ruido y valores faltantes

- Publicar los datos luego de procesos de revisión para reducir duplicados, errores de digitación o inconsistencias.
- Señalar si se ha realizado imputación, anonimización o normalización, y con qué criterios o herramientas.

4.3.5. Licencia compatible con IA

- Confirmar que el conjunto está licenciado bajo Creative Commons Atribución – Compartir Igual 4.0 Internacional (CC BY-SA 4.0), actualmente permitida en

www.datos.gov.co, y que permite usos comerciales, redistribución y adaptaciones, incluyendo desarrollos con IA.

- Incluir explícitamente en los metadatos que el conjunto puede ser utilizado para fines de innovación y desarrollo de modelos de inteligencia artificial.

4.4. Metadatos enriquecidos y estructuración semántica para IA

Los metadatos enriquecidos son elementos estructurados que describen, documentan y contextualizan los datos, permitiendo su descubrimiento, evaluación, interpretación y reutilización, especialmente en entornos automatizados como los sistemas de inteligencia artificial (IA). En el contexto de los datos abiertos, los metadatos no solo ayudan a identificar el contenido del conjunto de datos, sino que también actúan como facilitadores clave para la interoperabilidad semántica y el entrenamiento de modelos inteligentes.

Una adecuada gestión de metadatos mejora sustancialmente la calidad de los datos, al:

- Facilitar la búsqueda y comprensión del contenido publicado.
- Optimizar la interoperabilidad entre sistemas y plataformas.
- Permitir el trazado del origen, la autoría y los cambios de los datos.
- Habilitar procesos de verificación automatizados para IA.

Estándar DCAT: Un enfoque para la interoperabilidad

El portal de Datos Abiertos de Colombia adopta el estándar DCAT (Data Catalog Vocabulary), desarrollado por W3C, como marco para la definición de metadatos en sus catálogos. Este estándar permite:

- Integración semántica de datos entre diferentes catálogos y plataformas.
- Uso de vocabularios controlados y estructuras normalizadas.
- Alineación con buenas prácticas internacionales de apertura de datos.

¿Qué son los metadatos enriquecidos?

A diferencia de los metadatos mínimos, los metadatos enriquecidos incluyen campos adicionales que permiten ampliar el contexto y facilitar el procesamiento automatizado. Estos pueden incluir:

- Ontologías o vocabularios de dominio.
- Relación con otras fuentes de datos.
- Referencias temporales y espaciales más precisas.
- Enlaces a documentación técnica, códigos fuente, o API.

Buenas prácticas en la gestión de metadatos para IA

1. Utilizar títulos neutros y permanentes, evitando incluir fechas.
2. Redactar descripciones completas, que incluyan propósito, alcance y posibles limitaciones del conjunto.
3. Etiquetar con palabras clave relevantes para IA (por ejemplo: salud pública, clima, movilidad).
4. Incluir información clara sobre el formato, frecuencia de actualización y estructura del recurso.
5. Incorporar identificadores únicos persistentes (ej. URI) para facilitar la referenciación.
6. Documentar con precisión los campos y unidades de medida de cada variable.

Errores comunes que afectan la reutilización para IA

- Títulos ambiguos o con años desactualizados.
- Metadatos incompletos o vacíos (descripción, contacto, frecuencia, cobertura geográfica).
- Falta de estandarización en formatos (mezcla de nombres de columnas, estructuras inconsistentes).
- Ausencia de vocabularios controlados o inconsistencias semánticas.
- Licencias de uso no definidas o restrictivas.

Recursos y herramientas disponibles

El portal www.datos.gov.co ofrece herramientas para el diligenciamiento adecuado de los metadatos, entre ellas:

- Guías y videos tutoriales paso a paso.
- Cursos virtuales dirigidos a entidades y ciudadanía.
- Acceso a APIs de consumo e integración.
- Validadores de estructura y consistencia.

Cómo diligenciar los metadatos en datos.gov.co

Título y descripción: Evitar fechas, utilizar lenguaje claro y descriptivo.

Etiqueta de la fila: Incluir un identificador único por registro.

Categorías y etiquetas: Elegir dominios pertinentes y tags relacionados con el uso potencial del conjunto.

Licencia de atribución: Seleccionar Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0) según lo establecido en el portal.

Información de contacto: Usar correo institucional de la dependencia responsable.

Nombre del recurso y anexos: Identificar claramente cada archivo e incluir documentación técnica de apoyo.

Información institucional y de cobertura: Precisar la entidad, área responsable, alcance geográfico y temporal.

URL de documentación y normativa: Enlazar a guías, decretos, normas u otros documentos que contextualicen el conjunto de datos.

La calidad de los datos está directamente relacionada con la calidad de sus metadatos. Para que un conjunto de datos sea útil en entornos de inteligencia artificial, los metadatos deben ser completos, actualizados y estructurados con un enfoque interoperable, permitiendo su integración y análisis automatizado de manera efectiva y confiable.

4.5. Evaluación y mejora continua de los datos habilitantes para IA

La transformación digital del Estado y la incorporación de tecnologías de inteligencia artificial (IA) requieren que las entidades públicas implementen procesos sistemáticos para evaluar y mejorar continuamente la calidad de sus datos. Este enfoque no solo responde a una exigencia técnica, sino a un compromiso institucional con la generación de valor público, la toma de decisiones basada en evidencia y la confianza ciudadana.

Los conjuntos de datos habilitantes para IA —aquellos que permiten entrenar, evaluar o implementar modelos inteligentes— deben someterse a ciclos iterativos de revisión que garanticen su utilidad, confiabilidad, integridad y oportunidad.

4.5.1. Evaluación periódica basada en criterios de calidad

Cada conjunto de datos publicado debe ser evaluado de manera periódica bajo los criterios definidos en esta guía, tales como: exactitud, unicidad, completitud,

portabilidad, trazabilidad, recuperabilidad y relevancia. Estos criterios permiten identificar debilidades técnicas que, de no ser abordadas, pueden afectar la validez de los resultados en procesos automatizados o de IA.

Para facilitar este proceso, se recomienda el uso de herramientas automáticas de evaluación de calidad, cuyos lineamientos técnicos deben ser definidos por el MinTIC y sus entidades aliadas en la política de datos abiertos. La medición de calidad debe ser integrada a las actividades regulares de gestión de datos en las entidades.

4.5.2. Seguimiento de la reutilización y retroalimentación

Además de la revisión técnica, es fundamental evaluar cómo los conjuntos de datos son utilizados en la práctica. La recopilación de estadísticas de uso (como descargas, llamadas a la API, citas en investigaciones o reutilización en servicios públicos) permite priorizar esfuerzos de mejora sobre los conjuntos más demandados o con mayor potencial para habilitar soluciones de IA.

Asimismo, los canales de retroalimentación por parte de usuarios (ciudadanos, academia, empresas u otras entidades públicas) deben ser activamente promovidos y atendidos. Las observaciones sobre errores, ambigüedades o desactualizaciones en los conjuntos publicados son insumos valiosos para su mejora continua.

4.5.3. Planes de mejora institucional y roles definidos

La gestión de la calidad de los datos debe estar respaldada por planes de mejora con responsables, acciones y plazos definidos. Estos planes deben articularse con los procesos institucionales de planeación, tecnología, transparencia y participación ciudadana.

En este sentido, los líderes de tecnología (CIOs) y responsables de datos abiertos en cada entidad deben ejercer un rol estratégico, promoviendo una cultura de mejoramiento continuo basada en datos y facilitando la articulación con otras áreas como planeación, jurídica, comunicaciones y sectoriales.

5. Glosario

API (Interfaz de Programación de Aplicaciones): Conjunto de definiciones y protocolos que permiten que diferentes programas se comuniquen entre sí para consultar, enviar o recibir datos automáticamente.

Calidad de datos: Grado en que los datos son precisos, completos, consistentes, actuales, relevantes y comprensibles, adecuados para su propósito previsto.

Creative Commons (CC BY-SA 4.0): Licencia internacional que permite el uso, modificación y redistribución de los datos, siempre que se otorgue crédito al autor original y las nuevas obras se compartan bajo las mismas condiciones.

DCAT (Data Catalog Vocabulary): Especificación desarrollada por el W3C que permite describir catálogos de datos de manera estandarizada, facilitando su interoperabilidad entre sistemas.

Datos abiertos: Datos digitales que están disponibles de forma libre, sin restricciones técnicas o legales, para ser utilizados, reutilizados y redistribuidos por cualquier persona.

Interoperabilidad: Capacidad de diferentes sistemas y organizaciones para intercambiar datos y utilizar la información de manera efectiva.

Metadatos: Datos que describen otros datos. Incluyen información como el título, la descripción, la frecuencia de actualización, el formato, la cobertura geográfica y la licencia.

Metadatos enriquecidos: Conjunto ampliado de metadatos que incluye descripciones detalladas, relaciones semánticas, vocabularios controlados y otros elementos que facilitan su uso por herramientas automatizadas, especialmente para IA.

Ontología: Estructura formal que define conceptos y relaciones dentro de un dominio específico, facilitando la comprensión semántica de los datos por parte de máquinas.

Reutilización de datos: Proceso mediante el cual terceros acceden a conjuntos de datos abiertos y los emplean para desarrollar productos, investigaciones, análisis o servicios innovadores.

Trazabilidad: Capacidad de seguir el rastro de los cambios realizados a un conjunto de datos, desde su creación hasta sus actualizaciones más recientes.

6. Referencias

- Departamento Nacional de Planeación. (2025). **CONPES 4144 - Política Nacional de Inteligencia Artificial**. Recuperado de: <https://colaboracion.dnp.gov.co/CDT/Conpes/Econ%C3%B3micos/4144.pdf>
- Ministerio de Tecnologías de la Información y las Comunicaciones (MinTIC). (2024). **Guía para el uso y aprovechamiento de datos abiertos en Colombia**. Bogotá D.C. [https://drive.google.com/file/d/1jiuxTrsXyz_rEtKtpG9MR3qC-yTGUhw/view?usp=drive link](https://drive.google.com/file/d/1jiuxTrsXyz_rEtKtpG9MR3qC-yTGUhw/view?usp=drive_link)
- Open Knowledge Foundation. (2015). **Guía de Datos Abiertos**. Disponible en: <https://opendatahandbook.org>
- World Wide Web Consortium (W3C). (2020). **Data Catalog Vocabulary (DCAT) Version 2**. Disponible en: <https://www.w3.org/TR/vocab-dcat-2/>
- International Organization for Standardization (ISO). (2023). **UNE-ISO/IEC 25012:2023 – Requisitos de calidad de datos**.
- Portal de Datos Abiertos de Colombia. (2024). www.datos.gov.co – Términos y condiciones de uso, licenciamiento y herramientas de publicación.
- European Commission. (2017). **European Interoperability Framework (EIF)**. Bruselas: Comisión Europea.
- Cai, L., & Zhu, Y. (2015). **The Challenges of Data Quality and Data Quality Assessment in the Big Data Era**. *Data Science Journal*, 14(2), 1–10.
- Sunlight Foundation. (2017). **Principles for Open Data**. Disponible en: <https://sunlightfoundation.com/opendataguidelines/>

Anexo1. Errores Frecuentes de calidad y errores de publicación

La siguiente tabla presenta información clave sobre los tipos de errores comunes que se evalúan actualmente en los procesos de validación de conjuntos de datos. Este análisis permite a las entidades identificar las causas específicas detrás de rechazos, proporcionando información detallada que facilita la toma de decisiones estratégicas para optimizar los procesos de validación. Además, ofrece una guía clara para las entidades publicadoras al señalar qué aspectos deben corregirse, promoviendo la mejora continua y asegurando que las futuras cargas de datos cumplan con los estándares de calidad requeridos. Esta herramienta es fundamental para garantizar la precisión y confiabilidad de los datos, fortaleciendo la gestión pública y la transparencia en el uso de la información.

Código de Error	Categoría de Error	Tipo de Error	Descripción Error	Solución
ERR001	Metadata Errada, Incompleta y/o vacía.	Título y Descripción mal nombrados	El título y descripción del conjunto de datos está incompleto y no está generalizado, presenta siglas, caracteres especiales o información que no es clara para los usuarios del conjunto de datos. Tampoco debe contener vigencias.	Se recomienda actualizar el título y la descripción siguiendo las recomendaciones de calidad de las herramientas de apoyo
ERR002	Metadata Errada, Incompleta y/o vacía.	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
ERR003	Metadata Errada, Incompleta y/o vacía.	El campo de la metadata correspondiente al nombre de usuario debe vincular la entidad publicadora	El campo de la metadata correspondiente al nombre de usuario contiene un nombre particular de un funcionario se debe vincular el nombre de la entidad	Se recomienda con el fin de darle mayor reconocimiento a su entidad por favor actualizar la información de su usuario, en el campo "Nombre de usuario" en donde deberá diligenciar en nombre de la entidad
ERR004	Error conjunto de datos sin filas	Conjunto de datos no tiene filas con información	Actualmente su conjunto de datos no tiene registros, lo cual impide el máximo aprovechamiento de la	Se recomienda ELIMINAR el conjunto de datos y CREAR uno que contenga columnas y filas que complementen lo

			información por parte de los ciudadanos, por lo anterior, solicitamos por favor complementar la estructura y los datos para que genere mayor valor al ciudadano.	actualmente publicado o disponga la tabla en su sitio web. En caso de que la información solo sea de interés particular de la entidad se recomienda disponer esta información en su página web
ERR005	Error conjunto de datos con pocas filas	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. Debe ser mayor de 50 filas. (Esto no aplica para los conjuntos de datos obligatorios)	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.
ERR006	Error en conjunto de datos obligatorio (Ley 1712 de 2014 - Decreto 103 de 2015)	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.	No se está utilizando las plantillas de acuerdo con los articulados del Decreto 103 de 2015.	Se recomienda de acuerdo con lo establecido por la ley 1712 publicar como mínimo: REGISTRO DE ACTIVOS DE INFORMACIÓN, ÍNDICE DE INFORMACIÓN CLASIFICADA Y RESERVADA, ESQUEMA DE PUBLICACIÓN DE INFORMACIÓN La publicación de estos datos, independientemente del formato del archivo en el que se encuentren (Word, Excel, CSV), debe estar disponible de forma accesible y reutilizable.
ERR007	Error Filas, Columnas o Celdas vacías	Conjunto de datos con campos vacíos y/ o datos nulos	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.
ERR008	Error conjunto de datos con una sola columna	El conjunto de datos tiene una sola columna	Actualmente la estructura de su conjunto de datos cuenta con una única columna o un único campo	Se recomienda ELIMINAR el conjunto de datos y crear uno que contenga columnas y filas que complementen lo

			de datos, lo cual impide el máximo aprovechamiento del portal de datos abiertos.	actualmente publicado o disponga la tabla en su sitio web. Si la información a publicar es solo de interés particular de la entidad disponga la tabla en su sitio web.
ERR008_1	Error Conjunto de datos con pocas columnas	Conjunto de datos presenta muy pocas columnas	El conjunto de datos tiene muy pocas columnas (Menos de 5) 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
ERR008_2	Error Columnas mal nombradas o sin nombre	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.
ERR008_3	Error Columnas sin descripción en la metadata	Conjunto de datos presenta columnas sin descripción	El conjunto de datos presenta un error al nombrar una o varias columnas, no fue asignada una descripción a cada una de ellas.	Se recomienda hacer uso de herramientas de calidad de datos para estandarizar los campos e identificar los errores presentados.
ERR009	Error sin columnas de geolocalización	Error falta campo de geolocalización	El conjunto de datos presenta campo de dirección el cual no presenta estandarización, siempre que exista un campo de dirección es necesario incluir campos de geolocalización (latitud y longitud), con el fin de que los usuarios puedan reutilizar el conjunto de datos para realizar mapas de ubicación	Se recomienda incluir campos de geolocalización (latitud y longitud), con el fin de que los usuarios puedan reutilizar el conjunto de datos para realizar mapas de ubicación
ERR010	Enlace inválido	El conjunto de datos enlaza a una dirección a un archivo en formato inválido (PDF, JPG, PNG)	El enlace del conjunto de datos externos no permite la descarga directa de un conjunto de datos en formatos válidos: csv,	Se recomienda corregir el enlace del conjunto de datos externo, que permita la descarga directa de alguno de los formatos estándares abiertos soportados.

			<p>xls, xlsx, json, kml, kmz y zip (shapefile georreferenciado)</p> <p>NO es un dato abierto esta en un formato cerrado como PDF, .DOC, .PPT</p>	<p>Le recordamos que los datos abiertos son datos estructurados en filas y columnas, dispuestos en formatos abiertos como csv, xls, xlsx, json, kml, kmz y zip (shapefile georreferenciado) y deben cumplir con los principios de completitud, fuente primaria y debe presentar un alto nivel de desagregación</p>
ERR011	Completitud del Conjunto de datos	Conjunto de datos clasificado por periodos (Vigencias) / por tipologías	<p>Actualmente los conjuntos de datos están siendo publicados por periodos y/o algún tipo de clasificación, por lo cual se recomienda con el fin de que los datos abiertos cumplan con el principio de completitud</p>	<p>Se recomienda crear UN SOLO CONJUNTO DE DATOS e INCLUIR una columna de periodo(ejemplo vigencias, fechas(DD/MM/YYYY)) al conjunto de datos y se consoliden todos los conjuntos de datos en uno solo y se vaya incluyendo los registros a medida que se van generando nuevos datos en los diferentes años</p> <p>Este conjunto de datos es de gran importancia por lo cual recomendamos brindarle a los ciudadanos la información de manera consolidada que facilite el uso y análisis de esta información histórica y adicionalmente que garantice la actualización permanente del conjunto de datos consolidado.</p>
ERR012	Error conjunto de datos es VISTA / FILTRO de datos abiertos nacionales	Dato abierto es un VISTA / FILTRO de un conjunto de dato de una entidad nacional	<p>Este conjunto de datos es una VISTA / FILTRO de un dato abierto de una entidad nacional que consolida esta información.</p>	<p>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.</p>

ERR013	Publicación	El conjunto de datos esta mal estandarizado	El conjunto de datos presenta errores de calidad en la publicación	Se recomienda ELIMINAR el conjunto de datos y CREAR correctamente el conjunto
ERR014	Desactualizado (Fecha actualización de los datos)	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 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 esta
ERR015	Error conjunto de datos con estructuras similares	El conjunto de datos se asemeja a otro conjunto de datos creado por la entidad	La estructura de los conjuntos de datos es igual.	Se recomienda concatenar los conjuntos de datos.
ERR016	Datos agregados o totalizados	El conjunto de datos presenta agregaciones o totales	El conjunto de datos tiene registros agregados y totales, le recordamos que los datos abiertos deben publicarse en su máximo nivel de desagregación y completitud con el fin de maximizar el uso de los datos por parte de los ciudadanos	Se recomienda disponer la fuente de datos detallada con la cual se generó este reporte y hacer uso de la funcionalidad de vistas filtradas del portal para generar reportes
ERR017	Error falta de Comprensibilidad de la Historia con Datos	Conjunto de datos no permite ser interpretado de acuerdo con la historia con datos	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.

ERR018	Error Confidencialidad de los Datos	El conjunto de datos contiene datos personales	<p>El conjunto de datos incumple con el atributo de calidad de la confidencialidad</p> <p>Confidencialidad es el 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.</p>	Se recomienda ELIMINAR el conjunto de datos y aplicarle técnicas de anonimización de la información si requiere publicar con el fin de eliminar los datos personales del conjunto de datos
ERR019	Error en los tipos de Datos	Valores de los datos de acuerdo con sus formatos no pueden ser utilizados.	La estandarización de las columnas con el tipo de dato no es idónea de acuerdo con las variables del conjunto de datos.	Se recomienda estandarizar los tipos de datos de acuerdo con las columnas con las variables a segmentar.
ERR020	Error Falta Unicidad	Los registros o filas 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
ERR021	Error Poca reutilización - Uso y Aprovechamiento	La información contenida en el conjunto de datos no muestra interés a la Cuádruple Hélice (Academia, empresas, entidades públicas y ciudadanos)	El conjunto de datos puede ser no reutilizado por parte de la Cuádruple Hélice 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
ERR022	Error conjunto de datos es vista, filtro o grafica de	Dato abierto es un vista, filtro o grafica de un conjunto de	Este conjunto de datos es una vista, filtro o grafica de un conjunto de datos en	Se recomienda eliminar la vista, filtro o grafica del conjunto de datos privado / rechazado hasta que el

	conjunto de datos privado	datos en estado privado	estado privado / rechazado.	conjunto de datos sea aprobado.
--	---------------------------	-------------------------	-----------------------------	---------------------------------