



El futuro digital
es de todos

MinTIC

EJEMPLOS DE CRITERIOS DE CALIDAD

1. Criterio de confidencialidad



Este criterio establece que los datos solo deben ser accedidos por personas autorizadas para proteger adecuadamente la información reservada y clasificada.

Conjuntos de datos sin anonimizar:

Nombre	Género	Cédula	Número telefónico	Departamento	Enfermedades
Ricardo Rojas	Masculino	12345678910	3001234561	Antioquia	Cáncer
Felipe Agudelo	Masculino	12345678911	3001234562	Meta	Infección
Clara Ramírez	Femenino	12345678912	3001234563	Quindío	Renal
Estefanía López	Femenino	12345678913	3001234564	Antioquia	Ninguna
Camilo Zapata	Masculino	12345678914	3001234565	Quindío	Esclerosis
Carmen Vélez	Femenino	12345678915	3001234566	Meta	Sida
Sergio Rivas	Masculino	12345678916	3001234567	Antioquia	Ninguna
Santiago Álvarez	Masculino	12345678917	3001234568	Quindío	Cáncer
Ángela Ocampo	Femenino	12345678918	3001234569	Quindío	Listeria
Edison Martínez	Masculino	12345678919	3001234570	Antioquia	Sida
Ana Murillo	Femenino	12345678920	3001234571	Meta	Cáncer
Sofía Castaño	Femenino	12345678921	3001234572	Meta	Ninguna

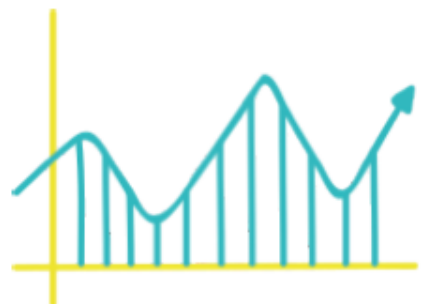


Conjunto de datos anonimizado

ID	Género	Departamento	Enfermedades
1	Masculino	Antioquia	Cáncer
2	Masculino	Meta	Infección
3	Femenino	Quindio	Renal
4	Femenino	Antioquia	Ninguna
5	Masculino	Quindio	Esclerosis
6	Femenino	Meta	Sida
7	Masculino	Antioquia	Ninguna
8	Masculino	Quindio	Cáncer
9	Femenino	Quindio	Listeria
10	Masculino	Antioquia	Sida
11	Femenino	Meta	Cáncer
12	Femenino	Meta	Ninguna

La Ley 1266 de 2008, también conocida como Ley de Habeas Data

2. Criterio de relevancia



Determina los temas importantes, destacados, significativos o con mayor demanda por parte de los usuarios que se deben publicar en la apertura de datos. Para esto, el Gobierno nacional define en su mapa de ruta algunas entidades, temáticas y criterios de priorización de acuerdo con la política de Gobierno Digital de MinTIC.

48 Entidades
16 categorías de información
93 conjuntos de datos priorizados

Temáticas priorizadas en el Mapa de Ruta

- Ordenamiento territorial
- Cadena productiva del agro
- Calidad y cobertura educativa
- Prestación de servicios de salud, salud pública y gestión del riesgo en salud
- Movilidad
- Seguridad ciudadana



El futuro digital
es de todos

MinTIC

3. Criterio de actualidad



La actualidad representa la vigencia y validez de los datos. Existen conjuntos de datos que se encuentran desactualizados en sus fechas, cifras periódicas o metadatos, por lo que es fundamental mantener la información actualizada constantemente.

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

- Hay datos que pierden su vigencia y utilidad rápidamente, por eso establezca un límite de publicación más ajustado según las capacidades de la institución.
- Defina un cronograma de publicación según la periodicidad que se estipulan en los metadatos con el fin de que el conjunto de datos se conserve actualizado.
- Implemente procesos de automatización de apertura de datos a través de ETL (Extracción, Transformación y Carga)



El futuro digital
es de todos

MinTIC

4. Criterio de trazabilidad



La trazabilidad permite determinar el flujo de procesos y acciones que se han generado sobre un conjunto de datos a través del tiempo. Este criterio es importante para la interoperabilidad de los conjuntos de datos entre las diferentes instituciones y para la reutilización de la información por parte de los usuarios.

Garantice la trazabilidad publicando todos los históricos de los conjuntos de datos que se tengan disponibles.



El futuro digital
es de todos

MinTIC

Resultados de la encuesta de software público a entidades - DAFP - 2017

Ciencia, Tecnología e Innovación

este conjunto de datos contiene los resultados de la encuesta de software público realizada a entidades públicas, la cual fue realizada por el Departamento Administrativo de la Función Pública (DAFP) durante el mes de octubre del año 2017...

Actualizado 26 de octubre de 2017
Datos proporcionados por Dirección de Gobierno Digital

Más

Sobre este conjunto de datos

Actualizado 26 de octubre de 2017	Información de la Entidad
Datos actualizados por última vez 26 de octubre de 2017	Municipio: Bogotá D.C.
Última actualización de metadatos 26 de octubre de 2017	Nombre de la Entidad: Ministerio de Tecnologías de La Información y Las Comunicaciones
Fecha de creación 23 de octubre de 2017	Orden: Nacional
	Sector: Ciencia, Tecnología e innovación
	Área o dependencia: Dirección de Gobierno Digital
	Departamento: Bogotá D.C.
Vistas 232	Información de Datos
Descargas 28	

5. Criterio de conformidad



La conformidad hace referencia al cumplimiento de lineamientos y estándares vigentes en el proceso de apertura de datos. En términos de calidad de datos, el criterio de conformidad está dividido en dos componentes: Metadatos y Lenguaje Común de Intercambio.

1. Metadatos

Los metadatos describen otros datos que explican y facilitan la información importante acerca de estos, lo que permite el acceso a los recursos de información y la reutilización de la misma.

2. Lenguaje Común de Intercambio de Información

El intercambio de información entre organizaciones puede realizarse a través de diferentes implementaciones técnicas de servicios de interoperabilidad.

La interoperabilidad se define como “capacidad de dos o más sistemas o componentes para intercambiar información y utilizar la información intercambiada.” (ISO/25010).

Según el Marco para la Interoperabilidad de la política de Gobierno Digital, esta se define como “el ejercicio de colaboración entre organizaciones para intercambiar información y conocimiento en el marco de sus procesos de negocio, con el propósito de facilitar la entrega de servicios en línea a ciudadanos, empresas y a otras entidades”.



El futuro digital
es de todos

MinTIC

6. Criterio de exactitud



Este criterio evalúa que los datos abiertos a los usuarios estén diligenciados correctamente y que su vez brinden información exacta en sus conjuntos de datos.

- **Datos numéricos:** los datos numéricos son aquellos que contienen números, dinero y porcentajes. Se utilizan para representar cantidades o medidas dentro de un conjunto de datos.

Correcto 🙌 1500,3
Incorrecto 🙌 1.500,3

- **Email:** todo conjunto de datos debe contener un email de la persona o la entidad encargada de subir la información en la web. Utilice uno de los siguientes formatos de entrada para ingresar los datos de correos electrónicos:

1. `Sam Gibson`
2. `sam.gibson@socrata.com`
3. `Sam Gibson <sam.gibson@socrata.com`

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

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

- **Ubicación geográfica:** se define utilizando la notación de grados decimales para latitud y longitud en el siguiente formato:

(xx.xxxx, -yy.yyyy) donde "xx.xxxx" es la latitud y "-yy.yyyy" es la longitud

- **Enlaces URL:** solo se aceptan tres formatos de URL: FTP, HTTP Y HTTPs. No se aceptan guiones bajos en las URL. Ingrese los enlaces a páginas web en uno de los siguientes formatos de entrada:

1. `Socrata`
2. `http://www.socrata.com/`
3. `Socrata (http://www.socrata.com)`

7. Criterio de completitud



El futuro digital
es de todos

MinTIC



La completitud hace referencia a datos completamente diligenciados. Para que un conjunto de datos cumpla a cabalidad con los criterios de calidad establecidos en la presente guía, es fundamental que los datos se presenten completos para el uso de los usuarios.

ID	codDep	nomDepartamento	Población2017	Población2018	Población2019	Población2020
1	91	Amazonas	77948	78830	79739	80682
2	5	Antioquia	6613118	6691030	6768388	6845093
3	88	Archipiélago de San An	77759	78413	79060	79693
...
30	76	Valle del Cauca	4708262			4853327
31	97	Vaupés	44500	44928	45367	45822

En los diccionarios de datos o etiquetas de campo tipo texto, incluya siempre opciones como **“No aplica”, “Otro”, “No sabe”, “No responde”, “No disponible”**, según se requiera.

8. Criterio de consistencia



Sus datos son consistentes cuando están libres de contradicción y son coherentes respecto a otros datos en el mismo contexto de uso. Esto quiere decir que se utilice la misma codificación de **variables y etiquetas** en todos los conjuntos de datos.

Ejemplo: si se utiliza codificación como 1: Mujer, 2: Hombre, para la variable "Sexo", utilizar siempre estas convenciones en todos los conjuntos de datos. Aplica también, y con mayor rigurosidad, para las codificaciones que están estandarizadas, como los códigos de la división político-administrativa del país.

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

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



9. Criterio de precisión



Los conjuntos de datos se deben publicar con el más alto nivel de desagregación posible, es decir, que se publiquen siempre desde la fuente primaria de la información.

Evite hacer agregaciones ya que puede causar que se sesguen algunas dimensiones asociadas.

ID	Barrio	Sexo	Asistentes
1	B1	Hombre	10
2	B1	Mujer	5
3	B2	Hombre	12
4	B2	Mujer	4
5	B3	Hombre	5
6	B3	Mujer	16

a) Fuente primaria

ID	Barrio	Asistentes
1	B1	15
2	B2	16
3	B3	21

b) Agregación por barrio

ID	Sexo	Asistentes
1	Hombre	27
2	Mujer	25

c) Agregación por sexo



10. Criterio de portabilidad



Se trata de la capacidad del conjunto de datos para ser transferido y adaptarse a diferentes formatos, es decir, ser reutilizado. Se refiere además a la facilidad con la que el conjunto de datos se puede procesar, acceder y utilizar fácilmente descargándolo o consumiéndolo a través de Interfaces de Programación de Aplicaciones. Para el caso del portal de datos abiertos de Colombia, se cuenta con la API de Socrata que soporta lenguajes de programación como Java, JavaScript, PHP, Ruby, Scala, Swift, .Net, entre otros.



Formato	Independiente de una aplicación	Formato estructurado de datos	Editabile	Observaciones
				No se recomienda para datos abiertos
			X	No se recomienda para datos abiertos
	X		X	Se puede usar para datos abiertos
		X	X	Se puede usar para datos abiertos
	X	X	X	Recomendable para datos abiertos
	X	X	X	Recomendable para datos abiertos
	X	X	X	Recomendable para datos abiertos

11. Criterio de credibilidad



La credibilidad da cuenta del grado en el cual el conjunto de datos posee información que pueda ser catalogada como veraz y confiable por los usuarios. Esto se puede estimar con la presencia de información en los metadatos sobre fuentes de información, documentación, normatividad, origen y/o entidad publicadora de los datos.



El futuro digital
es de todos

MinTIC

- Diligenciar rigurosamente los campos de los metadatos que dan cuenta sobre el origen o autor de los datos.
- Indicar un contacto en los metadatos (nombre y correo) el cual pueda resolver todas las inquietudes que los usuarios tengan sobre el conjunto de datos . Procurar establecer un correo general (no personal) que sea revisado constantemente.
- Crear los usuarios de carga de información de las diferentes instituciones dentro del portal de datos abiertos www.datos.gov.co. Estos deben utilizar una imagen institucional y crear nombres de usuario alusivos a la institución.

12. Criterio de comprensibilidad



Los datos deben poseer atributos que permitan ser leídos e interpretados por los usuarios. En los metadatos se evalúan atributos como el nombre del conjunto de datos, descripción, diccionario de datos e índice de etiquetas. En el conjunto de datos se evalúa la claridad en los nombres de los campos y las etiquetas.

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

P1: ¿cuántos hermanos tiene?
P2: ¿cuál es su ocupación?
P3: ¿cuál es su estado civil?



El futuro digital
es de todos

MinTIC

- Para etiquetas o valores de los campos codificados, incluya en los metadatos la definición de dichas etiquetas, ejemplo:



0: Hombre, 1: Mujer

1: Pequeño, 2: Mediano, 3: Grande

- Siempre incluya en los metadatos una descripción narrativa de lo que significa cada campo, ejemplo:

Población2018: se refiere a la cantidad de personas estimadas según las proyecciones de población para cada uno de los departamentos de Colombia en el año 2018.

- Incluya una descripción clara de lo que significa cada registro del conjunto de datos, ejemplo:

Cada registro (o fila) de este conjunto de datos representa un departamento de Colombia y el valor de su proyección poblacional para los próximos 3 años.

13. Criterio de accesibilidad



La accesibilidad es uno de los principios acordados en la carta internacional de datos abiertos la cual pretende facilitar la disposición, búsqueda y descarga de datos para cualquier persona mediante la disposición de los recursos tecnológicos necesarios para garantizarlo. Los datos deben ser fácilmente visibles y accesibles, poniéndose a disposición sin barreras

- Publicar los datos en un portal central para que los datos abiertos se puedan encontrar fácilmente y estén accesibles en un solo lugar.
- Liberar los datos en formatos abiertos con el fin de asegurar que estos estén disponibles para el más amplio rango de usuarios; que puedan encontrarse, accederse y utilizarse, proporcionando los datos en múltiples formatos estandarizados, de modo que puedan procesarse por computadoras y utilizarse por personas.
- Liberar los datos de manera gratuita, sujetos a una licencia abierta y sin restricciones.
- Liberar los datos sin registro obligatorio, permitiendo a los usuarios escoger y descargar los datos sin requerir que se identifiquen.
- Asegurar que los datos puedan ser accesibles y usados eficazmente por el más amplio rango de usuarios.



El futuro digital
es de todos

MinTIC

HASTA AQUÍ

ERRORES COMUNES

1. Que el conjunto de datos tenga menos de cincuenta registros (a menos de que la información sea la única existe y/o disponible).
2. Que el conjunto de datos tenga en sus metadatos campos incompletos y/o vacíos.
3. Que el conjunto de datos tenga registros en blanco.
4. Que se carguen archivos en formatos cerrados que no corresponden a conjuntos de datos tales como formatos .pdf, .doc, .gif, .jpg, .ppt, entre otros.
5. Enlaces a sistemas de información que no permitan la descarga directa de un conjunto de datos en formato abierto: csv, xls,xlsx, json, kml, kmz y zip (shapefile de ESRI).
6. Que el conjunto de datos contenga datos personales publicados sin contar con la autorización de los titulares de acuerdo con la normatividad vigente: Ley 1581 de 2012 – Protección de datos personales.
7. Que el conjunto de datos incluya información clasificada y/o reservada de acuerdo con lo establecido en la Ley 1712 de 2014 - Transparencia y acceso a la información pública.
8. Enlaces rotos de los metadatos y/o conjuntos de datos.
9. Que el conjunto de datos esté compuesto de una única columna.
10. Que el conjunto de datos no cuente con ninguna información en formato de columnas y/o filas.”