



GUÍA DE ESTÁNDARES DE CALIDAD E INTEROPERABILIDAD DE LOS DATOS ABIERTOS DEL GOBIERNO DE COLOMBIA

- Datos abiertos para la transparencia y el control social
- Datos abiertos para empoderar al ciudadano
- Datos abiertos para mejorar o crear productos y modelos de negocio
- Datos abiertos como una puerta abierta a la innovación
- Datos abiertos para mejorar la eficiencia y eficacia del Estado
- Datos abiertos para medir el impacto de las políticas



GOBIERNO DE COLOMBIA



MINTIC



INTRODUCCIÓN

En Colombia, se han llevado a cabo en los últimos años importantes esfuerzos para impulsar la apertura de datos públicos, los cuales han tenido resultados significativos como se puede evidenciar en la plataforma www.datos.gov.co.


Además de lograr que las instituciones publiquen los datos que producen como resultado de su trabajo, es necesario que esto se haga bajo estándares que garanticen la calidad y permitan la interoperabilidad de los datos. Lo anterior constituye un importante desafío.

El objetivo de la presente guía es proporcionar orientaciones y buenas prácticas para el desarrollo de estrategias de apertura de datos que cumplan con los principales estándares de calidad e interoperabilidad de datos abiertos de los países mejor calificados en índices internacionales como el Open Data Barometer, el Global Open Data Index y el Our Data Index. Además de lo establecido en el mapa de ruta para la apertura de datos abiertos del Gobierno de Colombia, en las guías del dominio de información del Marco de Referencia del Ministerio de Tecnologías de la Información y las Comunicaciones (MinTIC), el modelo de calidad de datos ISO 25012, la Carta Internacional de datos abiertos y el Marco de Interoperabilidad para el Gobierno Digital.

Con base en este análisis se definieron 16 criterios fundamentales para la apertura de datos de calidad por parte del Gobierno de Colombia, los cuales son: conformidad, exactitud, completitud, consistencia, credibilidad, trazabilidad, actualidad, relevancia, confidencialidad, precisión, comprensibilidad, accesibilidad, eficiencia, recuperabilidad, disponibilidad y portabilidad.

Los criterios enunciados están orientados a la estandarización del proceso de apertura de datos, la cual permite utilizar y reutilizar la información para que pueda ser interoperable por diferentes entidades a nivel nacional e internacional.

El proceso de apertura de datos por parte de las diferentes instituciones a nivel nacional, regional y local deberá contener los 16 criterios de calidad seleccionados a manera de estándar. Con la aplicación de estos se busca lograr mejores calificaciones en las evaluaciones nacionales e internacionales de calidad e interoperabilidad de datos, replicar las mejores prácticas de los países líderes en datos abiertos y contar con un modelo de calidad estándar como la ISO 25012, con el fin de generar gran valor y aprovechamiento por parte de los diferentes usuarios, así como una mejor gestión y gobernanza en los sistemas de información que los contienen, incluyendo la plataforma de datos abiertos de Colombia: www.datos.gov.co.



¿POR QUÉ ES IMPORTANTE PUBLICAR DATOS ABIERTOS DE CALIDAD?



En la década de 1950, los investigadores iniciaron estudios sobre la calidad de los productos, los cuales llevaron a definiciones de calidad como la acuñada por Crosby en 1988, “conformidad con los requisitos”.

Más tarde, durante la década de 1990, el rápido desarrollo de las tecnologías de la información condujo al estudio de la calidad de los datos y otorgó también varias definiciones al respecto. Entre estas se destaca la del grupo de gestión de calidad de datos total de la Universidad del MIT, liderado por Wang & Strong (1996): “aptitud para el uso”.

Por ese motivo, en los últimos años se han realizado grandes esfuerzos para la estandarización de los datos abiertos, teniendo como propósito la reutilización de la información a través de estándares y la interoperabilidad de los datos por parte de cualquier usuario.

A continuación, se presentan los criterios que debe contener un conjunto de datos para que pueda ser de calidad e interoperable y que son definidos por el estándar universal de la calidad de los datos de 2 capas.

Estándar universal de calidad de datos de dos capas

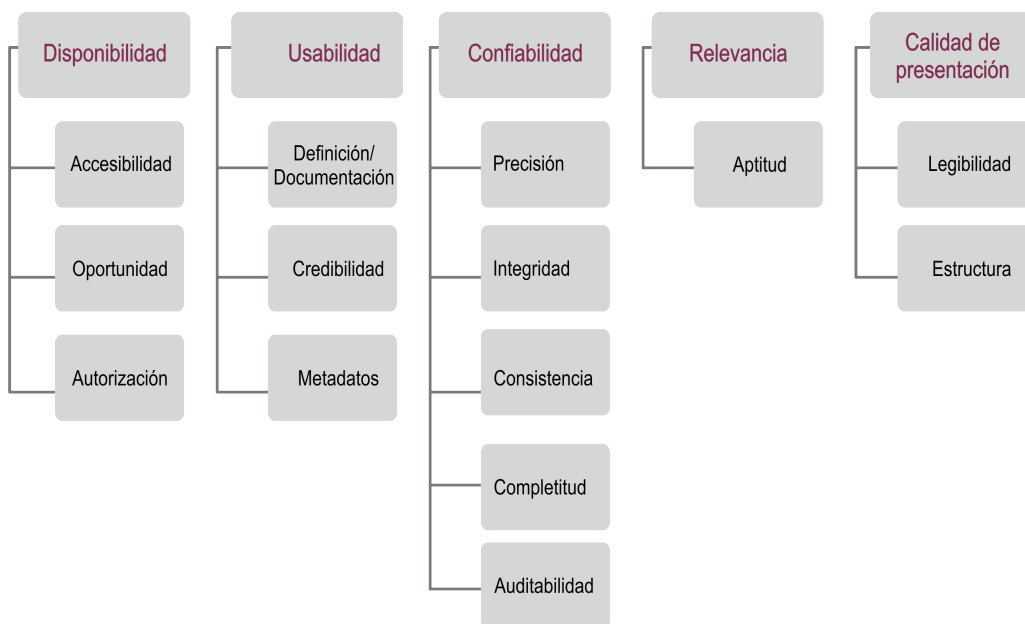


Imagen 1. Estándar universal de calidad de dos capas. Los desafíos de la calidad de los datos y la evaluación de la calidad de los datos en la era del Big Data. (Li CAI, Yangyong Zhu).[1]

Los datos abiertos de alta calidad son una condición previa para analizarlos, reutilizarlos y garantizar el valor de los datos. Actualmente, no todos los conjuntos de datos proporcionados y publicados por las entidades del Estado en el portal de datos abiertos de Colombia, www.datos.gov.co, tienen la misma calidad. Esta característica de los conjuntos de datos determina qué tanto se puede confiar en ellos para reutilizarlos y transformarlos en productos, servicios o nuevas aplicaciones. Por lo anterior, es importante evaluar la calidad de los datos para verificar su exactitud, confiabilidad y aptitud para el uso, es decir, la capacidad de los datos de prestar adecuadamente una función o servicio. En conclusión, los datos deben permitir transformarse, georreferenciar, analizar, reutilizar, visualizar y agregar sin necesidad de depurarse previamente.

[1] <https://datascience.codata.org/articles/10.5334/dsj-2015-002/>



¿CÓMO MEJORAR LA CALIDAD DE LOS DATOS?

La Guía de datos abiertos en Colombia precisa que una estrategia de apertura de datos va más allá de la preparación y publicación de los datos por parte de las entidades, también involucra acciones concretas para promover el uso de los datos publicados, realizar seguimiento e identificar el valor agregado que se está generando con su utilización. En la segunda fase del proceso de apertura y uso de los datos, “ Monitorear la calidad y el uso” , en la cual se centra esta guía, se propone analizar y presentar un paso a paso para el cumplimiento de los criterios definidos para estandarizar la calidad y mejorar la interoperabilidad de los datos abiertos publicados o por publicar en el portal correspondiente, los cuales se detallan en el capítulo **CRITERIOS DE CALIDAD E INTEROPERABILIDAD** de esta guía.

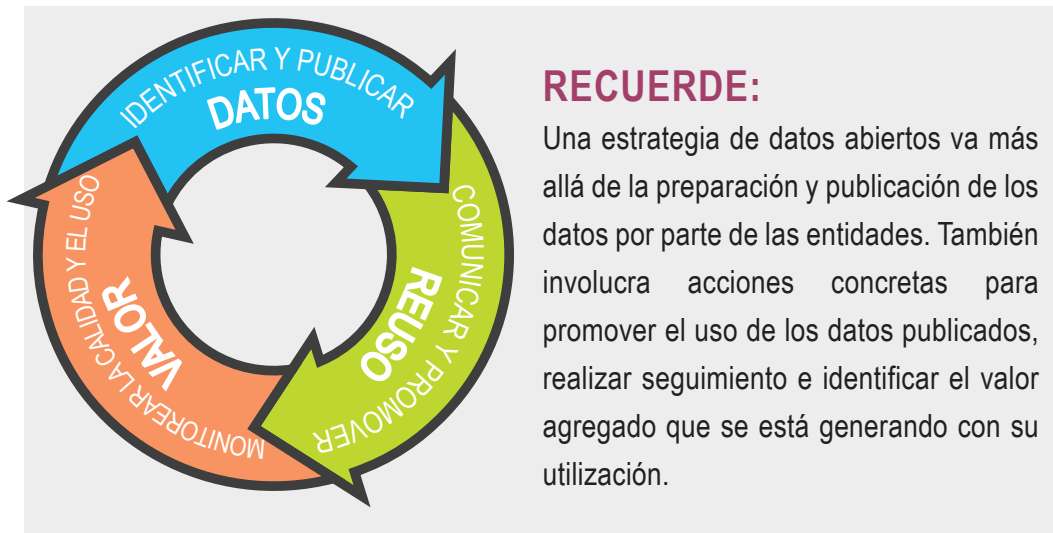


Imagen 2. Círculo virtuoso de los datos abiertos

CONOZCA EL MARCO PARA LA INTEROPERABILIDAD DEL GOBIERNO DIGITAL



En Colombia, el Marco para la Interoperabilidad del Gobierno Digital surge para contribuir en la garantía de la entrega de servicios e información del Estado, de manera completa y adecuada a la ciudadanía. El objetivo es minimizar pasos innecesarios y evitar que sean los usuarios quienes se tengan que desplazar a las entidades para obtener información útil sobre sus derechos y obligaciones. (Marco para la interoperabilidad del Gobierno Digital, 2010).

La política de interoperabilidad de MinTIC está orientada a:

- Entidades públicas responsables de planear servicios que requieran colaboración interinstitucional.
- Organizaciones privadas involucradas en la ejecución y/o evolución de la Política de Gobierno Digital.
- Miembros de gobiernos extranjeros interesados en la interoperabilidad del Gobierno Digital.
- Miembros de la comunidad académica interesados en la interoperabilidad del Gobierno Digital.

El Marco para la Interoperabilidad del Gobierno Digital comprende un conjunto de principios y políticas que orientan los esfuerzos políticos, legales y organizacionales de las entidades, con el fin de facilitar el intercambio seguro y eficiente de información.^[2]

[2] http://www.mintic.gov.co/arquitecturati/630/articles-9375_marco_interoperabilidad_pdf.pdf



El Marco de Interoperabilidad contempla cinco dominios:

Político-legal: corresponde a la disposición de un conjunto de

- políticas y normas que permiten el intercambio de información.

Sociocultural: hace referencia a la generación de competen-

- cias en las entidades para poder intercambiar información y a la habilitación de medios para la colaboración entre entidades.

Organizacional: modo en que las misiones, políticas, procesos

- de negocio y mecanismos de prestación de estos procesos de una entidad interactúan con aquellos de otras entidades a través del intercambio de información.

Semántico: permite garantizar que en el momento de intercam-

- biar datos, el significado de la información sea el mismo para todos los actores involucrados.

Técnico: hace referencia a las condiciones que se deben cum-

- plir para conectar los sistemas de información con el propósito de intercambiarla (Marco para la interoperabilidad del Gobierno Digital, 2010).

Con la adecuada implementación de los criterios de calidad desarrollados en esta guía, los dominios semánticos y técnicos se convierten en una consecuencia natural proveyendo las condiciones físicas necesarias y que servirán de base para la ejecución de los dominios político-legal, organizacional y sociocultural del marco de interoperabilidad.



CRITERIOS DE CALIDAD E INTEROPERABILIDAD

Parte 1.

1. Confidencialidad

Los datos solo deben ser accedidos por personas autorizadas para proteger adecuadamente la información reservada y clasificada.

2. Relevancia

Los datos publicados deben ser de utilidad para los usuarios.

3. Actualidad

Vigencia y actualidad de los datos publicados.

4. Trazabilidad

Histórico del conjunto de datos disponible: fechas de creación, publicación y actualizaciones.

5. Conformidad

Cumplimiento de lineamientos y estándares vigentes.

6. Exactitud

Datos diligenciados correctamente.

7. Completitud

Campos completamente diligenciados.

8. Consistencia

Datos coherentes y libres de contradicción.

9. Precisión

Nivel de desagregación en que están publicados los datos con respecto al nivel de desagregación en que fueron generados.



CRITERIOS DE CALIDAD E INTEROPERABILIDAD

Parte 2.

10. Portabilidad

Formatos sin restricciones para la reutilización de los datos.

11. Credibilidad

Información veraz y confiable para los usuarios.

12. Comprensibilidad

Características que permiten al usuario leer e interpretar los datos.

13. Accesibilidad

Herramientas tecnológicas que garanticen el acceso al usuario que lo requiera.

14. Eficiencia

Capacidad de la plataforma de realizar análisis y descargas de los datos con unos niveles de desempeño y tiempos esperados.

15. Recuperabilidad

Atributos que permiten mantener y preservar un nivel específico de operaciones y de calidad.

16. Disponibilidad

Garantiza que los usuarios autorizados tengan acceso a la información y a otros activos de información asociados en el lugar, momento y forma en que es requerido.

CRITERIOS DE CALIDAD E INTEROPERABILIDAD

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.

En Colombia, los criterios para asignar la calificación a las categorías de información son aquellos indicados por la Ley de Transparencia y del Derecho de Acceso a la Información (ley 1712 de 2014) la cual fue reglamentada con el Decreto 103 de 2015, en el que se establece que para la publicación de Datos Abiertos se debe tener en cuenta el título relativo a las “excepciones de acceso a la información”, el cual resalta la procedencia de limitar el acceso a la Información Pública Clasificada y la Información Pública Reservada.

A continuación, se describen los tipos de información que acota la ley y que deben ser tenidos en cuenta para validar el criterio de confidencialidad:



Imagen 3. Calificación de la información de acuerdo con sus niveles de seguridad [3]

El acceso a la información pública exige un ejercicio constante de ponderación entre el derecho a la información y el derecho a la protección de los datos personales.

En este contexto surge la anonimización como una herramienta para mitigar los riesgos que presentan la obtención y el tratamiento masivo de los datos públicos, el cual es un proceso que permite identificar y ocultar la información sensible contenida en los documentos, permitiendo su divulgación sin que ello implique vulnerar los derechos a la protección de datos de las personas y entidades que se puedan referenciar en los mismos. A continuación, se describen el concepto relacionado con la anonimización de la información:

Garantice que no sea posible identificar de manera directa o indirecta a individuos o entidades.

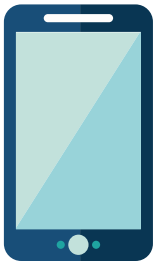
[3] <http://es.presidencia.gov.co/dapre/DocumentosSIGEPRE/G-GD-02-calificacion-informacion.pdf>

¿Qué es anonimizar la información?

El suministro de datos e información se convierte en un insumo esencial en la actualidad para minimizar el riesgo en la toma de decisiones asociadas a todas las esferas sociales y políticas a nivel mundial. El procesamiento y análisis de grandes volúmenes de información puede ofrecer cuantiosos beneficios a la sociedad, siempre y cuando se respete la privacidad de cada individuo o entidad y se protejan sus datos personales. Por tal motivo, se hace necesario establecer e implementar procesos de anonimización de los datos antes de ser disponibles al público en el portal de datos abiertos de Colombia.

Para que la anonimización sea exitosa, se debe garantizar que no sea posible identificar de manera directa o indirecta a individuos o entidades. Cabe resaltar que el avance tecnológico y la información disponible en distintos medios y en particular, a través de internet, dificulta un anonimato absoluto, principalmente en el transcurrir del tiempo. Aun así, la implementación de procesos de anonimización va ofrecer mayor seguridad en conservar el anonimato y la privacidad de las personas. Por tal motivo, es fundamental que las entidades realicen de una forma adecuada el proceso de anonimización pues, de no ser así, esto podría generar una reducción de la confianza de quienes gestionan la información y afectar considerablemente la calidad de los datos.

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



Pre-anonimización

Ningún conjunto de datos deberá permitir la identificación directa o indirecta de una persona. Por tal motivo, es recomendable eliminar la información personal con la que se va a trabajar y tener un especial cuidado con aquellos datos que contienen información sensible según la normatividad colombiana.



Ocultamiento, supresión y pseudo-anonimización

Identifique en su conjunto de datos aquellos datos que sean sensibles para su eliminación o sustitución. A continuación, algunos ejemplos de datos sensibles:

- Nombres.
- Fecha de nacimiento y fecha de constitución en Cámara de Comercio, (excepto el año).
- Números de teléfono y fax.
- Números de identificación: cédula de ciudadanía, pasaporte y tarjeta de identidad.
- Números asociados a la seguridad social, licencias de conducción, Número de Identificación Tributaria (NIT), Registro Único Tributario (RUT), Registro Único de Proponentes (RUP) y Registro Único Empresarial (RUES).
- Direcciones de correo electrónico.
- Números de cuentas bancarias.
- Identificadores del vehículo como placa, entre otros.
- Identificadores de dispositivos móviles y números de serie.
- Direcciones de IP.
- Cualquier otro número único de identificación.
- Dirección de domicilio.

Una vez identificadas estas variables, proceda a eliminar los campos del conjunto de datos o reemplácelos por algún seudónimo.

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

Imagen 4. Conjuntos de datos datos sin anonimizar.

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

Imagen 5. Conjuntos de datos anonimizados.

Para bases de datos extensas y masivas, estos procesos primarios suelen ser suficientes para anonimizar los conjuntos de datos.

Generalización

Evite el intento de identificación de una persona a través de sus datos.

Generalice los atributos a través de la modificación de escalas o magnitudes (por ejemplo, cambiando una ciudad o municipio por un departamento, o días por meses). También puede ser aplicada creando rangos en lugar de valores específicos (por ejemplo, cambiar la edad absoluta por un rango de edades o grupo etario). En la imagen 6 se presenta un ejemplo de un conjunto de datos con información de edad e ingresos absolutos, lo cual podría ser utilizado para singularizar un individuo. En la imagen 7 se muestra el mismo conjunto de datos después de generalizar la edad y los ingresos, estableciendo rangos que dificultan un intento de identificación.

ID	Edad	Enfermedades	Departamento	Salario
1	18	Cáncer	Antioquia	1.500.000
2	22	Infección	Meta	9.000.000
3	64	Renal	Quindío	900.000
4	44	Ninguna	Antioquia	500.000
5	33	Esclerosis	Quindío	3.500.000
6	19	Sida	Meta	150.000
7	36	Ninguna	Antioquia	2.000.000
8	24	Cáncer	Quindío	1.500.000
9	55	Listeria	Quindío	850.000
10	39	Sida	Antioquia	1.000.000
11	68	Cáncer	Meta	1.200.000
12	15	Ninguna	Meta	5.000.000

Imagen 6. Conjunto de datos sin anonimizar.

ID	Edad	Enfermedades	Departamento	Salario
1	0-20	Cáncer	Antioquia	Entre 1 y 2 SMMLV
2	21-40	Infección	Meta	Más de 2 SMMLV
3	61-80	Renal	Quindío	Menos 1 SMMLV
4	41-60	Ninguna	Antioquia	Menos 1 SMMLV
5	21-40	Esclerosis	Quindío	Más 2 SMMLV
6	0-20	Sida	Meta	Menos 1 SMMLV
7	21-40	Ninguna	Antioquia	Más 2 SMMLV
8	21-40	Cáncer	Quindío	Más 2 SMMLV
9	41-60	Listeria	Quindío	Entre 1 y 2SMMLV
10	21-40	Sida	Antioquia	Entre 1 y 2SMMLV
11	61-80	Cáncer	Meta	Entre 1 y 2SMMLV
12	0-20	Ninguna	Meta	Más 2 SMMLV

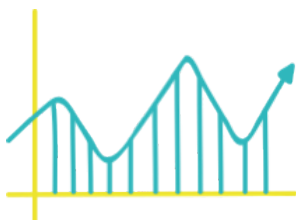
Imagen 7. Conjunto de datos anonimizado por supresión y pseudo-anonimización

Existen otras técnicas de anonimización descritas por el DANE para la producción de datos estadísticos que pueden ser consultadas en el siguiente enlace:

https://www.dane.gov.co/files/sen/lineamientos/DSO_020_LIN_08.pdf

CRITERIOS DE CALIDAD E INTEROPERABILIDAD

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.^[4]

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

Algunos índices internacionales como el Open Data Barometer, el Global Open Data Index y el Our Data Index definen unas temáticas a priorizar, las cuales son:

Temáticas priorizadas en el Open Data Barometer ^[5]

- Datos geográficos
- Datos de propiedad de la tierra
- Microdatos de censos
- Presupuesto gubernamental detallado
- Gastos gubernamentales
- Registro de la compañía
- Legislación

[4] <http://estrategia.gobiernoenlinea.gov.co/623/w3-article-9404.html>

[5] <http://opendatabarometer.org>



- Horarios de transporte público
- Comercio internacional
- Salud
- Educación primaria o secundaria
- Estadísticas de criminalidad
- Estadísticas del entorno nacional
- Resultados electorales
- Contratación pública

Temáticas priorizadas en el Global Open Data Index [6]

- Presupuesto
- Gastos
- Adquisiciones
- Resultados electorales
- Registro de la compañía
- Propiedad de la tierra
- Mapas nacionales
- Límites administrativos
- Ubicaciones
- Estadísticas nacionales
- Proyecto de legislación
- Ley nacional
- Calidad del aire
- Calidad del agua
- Predicciones meteorológicas (Discuss.okfn.org, s.f.)

Temáticas priorizadas en el Our Data Index [7]

- Negocios
- Registros
- Patentes y marcas
- Licitaciones públicas
- Información geográfica
- Legal
- Meteorología
- Datos Sociales
- Transporte

[6] <https://index.okfn.org>

[7] <http://dx.doi.org/10.1787/6f306823-en>

CRITERIOS DE CALIDAD E INTEROPERABILIDAD

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.

Aplique la prueba de actualidad a la última versión del conjunto de datos, pues es importante que permanezcan publicados los históricos de datos que son de gran importancia para la trazabilidad y los análisis temporales. Evalúe, además, que sus datos estén siendo publicados oportunamente en relación con la fecha en la que fueron producidos.

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

- Hay datos que pierden su vigencia y utilidad rápidamente, por eso establezca un límite de publicación más ajustado según las capacidades de la 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),^[8] para ello revise el Manual de Creación de ETL.

[8] http://herramientas.datos.gov.co/sites/default/files/CO_417_MANUAL_ETL.pdf

CRITERIOS DE CALIDAD E INTEROPERABILIDAD

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.

A manera de ejemplo, se recomienda consolidar todos los datos dispersos que están agrupados por vigencias o años en un solo conjunto de datos en donde se incluya una columna con el año de publicación de dichos datos.

Ejemplo: conjunto de datos del SECOP I, el cual consolida históricamente todos los procesos de compra pública del Estado.

Adicionalmente, la plataforma de datos abiertos ofrece la posibilidad de conocer las últimas fechas de actualización tanto de los datos, como de los metadatos, lo que permite conocer si están actualizados de acuerdo con la frecuencia definida para cada conjunto.

A continuación, la imagen 8 presenta un pantallazo de los metadatos de un conjunto de datos donde se evidencian las fechas de las últimas actualizaciones y la fecha de creación del conjunto, lo que permite un mayor aprovechamiento y seguimiento a la trazabilidad de los datos.

DATOS ABIERTOS
GOBIERNO DIGITAL COLOMBIA

Inicio Descubre Publica Visualiza Participa Herramientas Novedades Español Iniciar sesión

Resultados de la encuesta de software público a entidades - DAFF - 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
Vistas 232	Sector Ciencia, Tecnología e innovación
Descargas 28	Área o dependencia Dirección de Gobierno Digital
	Departamento Bogotá D.C.
	Información de Datos

Imagen 8. Trazabilidad de los conjuntos de datos en el portal de datos abiertos www.datos.gov.co.

CRITERIOS DE CALIDAD E INTEROPERABILIDAD

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.

Una buena gestión de los metadatos debe conducir a una buena calidad de los datos, ya que tenerlos y confiar en ellos puede ayudar a identificar información deficiente, incorrecta y faltante.

El portal de datos abiertos de Colombia adoptó el estándar DCAT para la definición de los metadatos de los conjuntos de datos [9], cuyo uso básico es hacer posible la interoperabilidad entre conjuntos de datos albergados en diferentes portales de datos abiertos.

[9] http://herramientas.datos.gov.co/sites/default/files/articles-8248_Guia_Apertura_Datos_0.pdf

En la siguiente imagen se detallan los metadatos requeridos durante el proceso de publicación de datos abiertos en www.datos.gov.co:

Crear un nuevo conjunto de datos

Por favor, describa sus datos.

* Título de Conjunto de Datos	<input type="text"/>
Breve Descripción	<input type="text"/>
Categoría	<input type="text" value="-- Sin categoría --"/>
Etiquetas/Palabras clave	<input type="text"/>
Título de Fila	<input type="text"/>
<p>Licencia y Atribución</p>	
Tipo de Licencia	<input type="text" value="-- Sin licencia --"/>
Datos Ofrecidos Por	<input type="text"/>
Enlace de la Fuente	<input type="text"/>

Imagen 9. Metadatos del conjunto de datos.

Asegúrese de que sus datos estén protegidos con alguna de las licencias para el uso e intercambio de información. Se recomienda proteger los conjuntos de datos a través de licencias abiertas estándar. Esta permite proteger la estructura de los conjuntos de datos, sus contenidos y garantizar que los usuarios de los datos atribuyan la titularidad de sus derechos a los creadores cuando se haga reuso de los mismos. [10]

La licencia Creative Commons, recomendada en los términos y condiciones de uso del portal de datos abiertos www.datos.gov.co, es **Atribución – Compartir igual (Attribution-ShareAlike 4.0 International)** CC BY-SA 4.0.



Atribución – Compartir igual: Esta licencia permite a otros remezclar, retocar, y crear a partir de tu obra, incluso con fines comerciales, siempre y cuando te den crédito y licencien sus nuevas creaciones bajo las mismas condiciones. Esta licencia suele ser comparada con las licencias “copyleft” de software libre y de código abierto. Todas las nuevas obras basadas en la tuya portarán la misma licencia, así que cualesquiera obras derivadas permitirán también uso comercial. Esa es la licencia que usa Wikipedia, y se recomienda para materiales que se beneficiarían de incorporar contenido de Wikipedia y/o proyectos con licencias similares.

Imagen 10. Licencia Creative Commons Atribución- Compartir igual Attribution-ShareAlike 4.0 International) CC BY-SA 4.0.[11]

[10] http://herramientas.datos.gov.co/sites/default/files/articles-8248_Guia_Apertura_Datos_0.pdf

[11] https://co.creativecommons.org/?page_id=13

Para acceder a más información acerca del diligenciamiento de los metadatos, remítase al portal de datos abiertos de Colombia: www.datos.gov.co, en la sección de herramientas y observe los videos sobre cómo publicar conjuntos de datos:

<http://herramientas.datos.gov.co/es/content/publicar-conjuntos-de-datos>.

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”.

A continuación, se presenta un ejemplo de cómo se debe realizar el proceso de armonización semántica sobre un conjunto de datos a publicar en formato abierto, utilizando la herramienta del Lenguaje Común de Intercambio de Información.

1. Seleccione el conjunto de datos a estandarizar.

Fecha de registro	Departamento	Ocupación
23 febrero 2017	Antioquia	Agricultor
15 marzo 2015	Cundinamarca	Oficial de construcción
10 diciembre 2016	Valle del Cauca	Vendedor

Imagen 11. Ejemplo de conjunto de datos sin utilizar los elementos del Lenguaje Común de Intercambio de Información

2. Ingresar a la página <http://lenguaje.mintic.gov.co/diccionario-de-elementos-de-datos> donde se encontrará el diccionario de elementos de datos.

3. Hacer la búsqueda de cada una de las variables del conjunto de datos que se pretende publicar e ingresar al elemento que mejor se adapte a la variable que se está usando. Hacer clic en el ícono de la columna “Acción”.

Consultar elementos de dato

Información General

Buscar Elemento de dato

ocupación

Busqueda Avanzada

BUSCAR

Resultado de la Búsqueda

Elementos Encontrados: 81

Elemento	Tipo de Elemento	Identificador	Capa	Proyecto	Módulo	Área	Acción
Ocupación	COMPUESTO	ocupacion	USO COMÚN	N/A	N/A	PERSONAL	
Actividad Ocupacional	COMPUESTO	actividadOcupacional	USO PROYECTOS	SecretariaSalud	ICNF	N/A	
Código Actividad Ocupacional	SIMPLE	codActividadOcupacional	USO PROYECTOS	SecretariaSalud	ICNF	N/A	
Código Grupo Primario Ocupación	SIMPLE	codGrupoPrimarioOcupacion	USO COMÚN	N/A	N/A	PERSONAL	
Código Ocupación	SIMPLE	codOcupacion	USO COMÚN	N/A	N/A	PERSONAL	
Código Posición Ocupacional	SIMPLE	codPosicionOcupacional	USO PROYECTOS	SecretariaSalud	ICNF	N/A	
Código Subgrupo Ocupación	SIMPLE	codSubgrupoOcupacion	USO COMÚN	N/A	N/A	PERSONAL	

Imagen 12. Diccionario de elementos de datos

4. Se abrirá entonces una ficha con una descripción más detallada del elemento, la entidad autora de este, su uso, formato y demás características. En este caso, la variable “Ocupación” es un elemento compuesto, por lo cual se deberán usar 2 elementos de dato simples para representarla: el código y el nombre de la ocupación, según se observa en el campo “Formato”. Nuevamente, hay que hacer clic sobre el ícono de la columna “Acción” de alguno de los elementos de dato simple para acceder a información más detallada.

5. Después de hacer clic e ingresar a la columna “Acción”, así deberá quedar la variable “Ocupación” con los 2 elementos de datos simples para que esta pueda ser interoperable. A continuación, se muestran cómo quedarían los elementos “nom Ocupación ” y “codOcupación” en el conjuntos de datos, los cuales representan los nombres y el código de ocupación aceptados e interoperables.

Valores Permitidos	
nomOcupacion	codOcupacion
FUERZA PÚBLICA	0
MIEMBROS DEL PODER EJECUTIVO, DE LOS CUERPOS LEGISLATIVOS Y PERSONAL DIRECTIVO DE LA ADMINISTRACIÓN PÚBLICA Y DE EMPRESAS	1
PROFESIONALES UNIVERSITARIOS, CIENTÍFICOS E INTELLECTUALES	2
TÉCNICOS, POSTSECUNDARIOS NO UNIVERSITARIOS Y ASISTENTES	3
EMPLEADOS DE OFICINA	4
TRABAJADORES DE LOS SERVICIOS Y VENDEDORES	5
AGRICULTORES, TRABAJADORES Y OBREROS AGROPECUARIOS, FORESTALES Y PESQUEROS	6
OFICIALES, OPERARIOS, ARTESANOS Y TRABAJADORES DE LA INDUSTRIA MANUFACTURERA, DE LA CONSTRUCCIÓN Y DE LA MINERÍA	7
OPERADORES DE INSTALACIONES, DE MÁQUINAS Y ENSAMBLADORES	8
TRABAJADORES NO CALIFICADOS	9

Imagen 13. Ejemplo de los 2 nuevos elementos, nomOcupacion, codOcupacion, del conjunto de datos.

6. Finalmente, así debería quedar el nuevo conjunto de datos con los nuevos nombres de las variables y/o los elementos de datos interoperables y su respectivo contenido según el ejemplo propuesto con la variable “Ocupación”.

Concepto	Valor
Tipo de Dato:	NUMÉRICO
Longitud Mínima:	N/A
Longitud Máxima:	1
N° Máximo de Decimales:	N/A
Valor Mínimo Permitido:	0
Valor Máximo Permitido:	9
Formato (Máscara):	9

Validación:
N/D

Correlación		
Elemento de Dato	Identificador	Tipo Correlación
Nombre Ocupación	nomOcupacion	DEPENDENCIA

Imagen 14. Conjunto de datos con los nuevos nombres de las variables y/o los elementos de datos interoperables según el ejemplo propuesto.

Para ampliar información sobre el Marco de Interoperabilidad, consulte

http://www.mintic.gov.co/arquitecturati/630/articles-9375_marco_interoperabilidad_pdf.pdf

CRITERIOS DE CALIDAD E INTEROPERABILIDAD

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.

Asegúrese de que los datos estén diligenciados correctamente para que cumplan con el criterio de exactitud. A continuación, se describen los tipos de datos soportados actualmente por el portal de datos abiertos de Colombia. [12]:

- **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

Imagen 15. Datos numéricos.

- **Fecha y hora:** los campos de fecha y hora deben de ser diligenciados mediante el formato de la norma ISO 8601.[13] Diligéncielos teniendo en cuenta los lineamientos.

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

[13] <https://www.iso.org/iso-8601-date-and-time-format.html>


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

Imagen 16. Formato de fechas básicas.

● **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`)

● **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`

● **Lógico/ binario:** se utiliza para representar valores binarios y lógicos, tales como 1/0 y Verdadero/Falso. En este campo tenga en cuenta 2 posibles valores que serán representados por uno y cero (1 y 0).

Valor binario	Etiqueta 1	Etiqueta 2	Etiqueta 3	Etiqueta 4
1	Si	Verdadero	Positivo	Aumento
0	No	Falso	Negativo	Disminución

Imagen 17. Uso del tipo de campo Lógico/ binario

Tenga en cuenta las siguientes consideraciones a la hora de utilizar los diferentes tipos de datos:

- No utilice más de un tipo de dato en la misma columna.
- En los campos numéricos, adopte codificación numérica para los casos especiales, por ejemplo: -88: “No aplica”, -89: “No disponible”, -98: “No sabe”, -99: “No responde”, -77: “Otro”. Esto con el fin de no combinar tipo de datos de texto con datos numéricos.
- Si el contenido de un campo depende de una etiqueta recurrente, asegúrese de que esta se utilice de manera uniforme en todo el conjunto de datos. Por ejemplo, si se requiere usar la etiqueta SI/NO en un campo, no mezclar etiquetas como Sí, sí, No, no.

CRITERIOS DE CALIDAD E INTEROPERABILIDAD

7. Criterio de completitud



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.

Recuerde que, de acuerdo con las Políticas de Privacidad y Condiciones de Uso, “la Dirección de Gobierno Digital podrá eliminar o cambiar los permisos de privacidad de los conjuntos de datos que presenten algunos de los siguientes errores:

1. Que el conjunto de datos tenga menos de cincuenta registros (a menos de que la información sea la única existente 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.”[14]

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

- En los diccionarios de datos o etiquetas de campo tipo texto, incluya siempre opciones como “No aplica”, “Otro”, “No sabe”, “No responde”, “No disponible”, según se requiera. Esto con el fin de no dejar espacios vacíos cuando no se tenga la información.
- No publique conjuntos de datos que contengan una sola columna.
- No publique conjuntos de datos incompletos.
- Si existen datos sobre un mismo tema distribuidos en diferentes fuentes de información, consolide un único conjunto de datos que represente la información lo más completa posible para los usuarios.
- Cuando los conjuntos de datos provienen de un modelo de datos relacional, tenga en cuenta que cada tabla no es un conjunto de datos.

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

Imagen 18. Error de completitud.

[14] <http://herramientas.datos.gov.co/es/terms-and-conditions-es>

CRITERIOS DE CALIDAD E INTEROPERABILIDAD

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

Imagen 19. Ejemplo de conjunto de datos con errores de consistencia

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

Imagen 20. Ejemplo de conjunto de datos después de corregir errores de consistencia

CRITERIOS DE CALIDAD E INTEROPERABILIDAD

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.

En el ejemplo de la imagen 15 se puede observar la fuente primaria (a) del registro de la asistencia a un evento desagregado por sexo y barrio. Si se hace una agregación por barrio (b) se pierde toda la información relacionada a sexo e, igualmente, si se hace una agregación por sexo (c) se pierde toda la información relacionada con barrio.

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

Imagen 21. Error de precisión.

CRITERIOS DE CALIDAD E INTEROPERABILIDAD

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. [15]

A continuación, se describen los formatos que actualmente soporta el portal de datos abiertos de Colombia, www.datos.gov.co para su descarga. Entre ellos se encuentran CSV, CSV for Excel, CSV for Excel (Europe), JSON, RDF, RSS, TSV for Excel y XML.

[15] <https://dev.socrata.com/libraries/>




Formato	Independiente de una aplicación	Formato estructurado de datos	Editable	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

Imagen 22. Formatos para la publicación de datos abiertos.

CRITERIOS DE CALIDAD E INTEROPERABILIDAD

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.

También, en el catálogo de datos se evalúa la credibilidad en la adecuada creación de los usuarios que representan a las diferentes entidades del Estado que publican los datos dentro de la plataforma. (ISO/IEC 25012). [16]

Tenga en cuenta que para garantizar el criterio de credibilidad debe:

- 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.

[16] <http://iso25000.com/index.php/normas-iso-25000/iso-25012?limit=5&start=5>

CRITERIOS DE CALIDAD E INTEROPERABILIDAD

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.

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

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

P1: ¿cuántos hermanos tiene?

P2: ¿cuál es su ocupación?

P3: ¿cuál es su estado civil?

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



Imagen 23. Etiquetas o valores de campos codificados.

- 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.

CRITERIOS DE CALIDAD E INTEROPERABILIDAD QUE GARANTIZA LA PLATAFORMA

13. Criterio de accesibilidad

Los criterios de accesibilidad, eficiencia, recuperabilidad y disponibilidad son fundamentales para la calidad e interoperabilidad de los datos. Estos criterios son inherentes al funcionamiento de la plataforma de datos abiertos

www.datos.gov.co



La accesibilidad es uno de los principios acordados en la carta internacional de datos abiertos [17], 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 burocráticas o administrativas que pueden disuadir a las personas de acceder a los datos. Para ello la plataforma de datos abiertos de Colombia permite:

- a. Publicar los datos en un portal central para que los datos abiertos se puedan encontrar fácilmente y estén accesibles en un solo lugar.
- b. 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.

[17] <https://opendatacharter.net/principles-es/>

- c. Liberar los datos de manera gratuita, sujetos a una licencia abierta y sin restricciones.
- d. Liberar los datos sin registro obligatorio, permitiendo a los usuarios escoger y descargar los datos sin requerir que se identifiquen.
- e. Asegurar que los datos puedan ser accesibles y usados eficazmente por el más amplio rango de usuarios.

Adicionalmente, se han creado iniciativas como Convertic, que promueve la inclusión y la autonomía de 1.2 millones [18] de personas con discapacidad visual en Colombia gracias al uso y apropiación de las TIC. Por medio de este proyecto, se ofrece a nivel nacional la descarga gratuita del lector de pantalla JAWS y el magnificador MAGIC.

JAWS: convierte a voz la información que se muestra en la pantalla, lo que permite a las personas con limitaciones visuales hacer un uso autónomo del computador y sus aplicaciones.

MAGIC: amplía hasta 16 veces el tamaño de las letras en pantalla y permite variar color y contraste, lo que beneficia a personas con baja visión o que estén empezando a experimentar problemas visuales.

Imagen 24. Herramientas JAWS y MAGIC de la iniciativa Convertic.

[18] <http://www.convertic.gov.co/641/w3-propertyvalue-15305.html>

CRITERIOS DE CALIDAD E INTEROPERABILIDAD QUE GARANTIZA LA PLATAFORMA

14. Criterio de eficiencia



Este criterio está relacionado con la plataforma y su capacidad de análisis y descargas de los datos con unos niveles de desempeño y tiempos esperados.

La actual plataforma de datos abiertos garantiza que los datos puedan ser procesados y proporcionados con los niveles de rendimiento esperados. Además, cuenta con protocolos que permiten una fácil conexión de los datos dispuestos en la plataforma para ser utilizados en otras aplicaciones directamente y en tiempo real, como pueden ser las herramientas de Business Intelligence, analítica e incluso Excel.

CRITERIOS DE CALIDAD E INTEROPERABILIDAD QUE GARANTIZA LA PLATAFORMA

15. Criterio de recuperabilidad



Programas, software, plataformas digitales y aplicaciones que permiten mantener y preservar un nivel específico de operaciones y calidad, incluso en caso de fallos, manteniendo la **recuperabilidad** de los datos.

El portal de datos abiertos de Colombia se reserva el derecho de remover el contenido en cualquier momento sin previo aviso si este está relacionado con difamación, calumnia, injuria, falsedad, obscenidad y pornografía dentro del servicio.

Por lo tanto, es importante que cada entidad tenga copias de seguridad de toda la información subida al portal y no lo tome como la fuente primaria de almacenamiento.

CRITERIOS DE CALIDAD E INTEROPERABILIDAD QUE GARANTIZA LA PLATAFORMA

16. Criterio de disponibilidad



Este criterio de calidad garantiza que los usuarios autorizados tengan acceso a la información y a otros activos asociados en el lugar, momento y forma en que es requerido. Una plataforma segura debe mantener la información disponible para los usuarios.



La plataforma de datos abiertos de Colombia cuenta con herramientas que garantizan que los sistemas de información de las entidades estatales puedan integrarse automatizando la publicación de sus datos en www.datos.gov.co, a través de procesos de Extracción, Transformación y Carga (ETL)^[19]. De igual manera, permite la consolidación e integración de datos de otros portales a través de procesos como la federación.^[20]

Para proveer la disponibilidad requerida de los datos abiertos, la plataforma utiliza estándares de conectividad y dispone de interfaces, tales como:

ODATA: es un estándar que permite el consumo de datos de manera fácil y en tiempo real a través de herramientas de Business Intelligence (BI), como por ejemplo Tableau o Power BI.^[21]

[19] http://herramientas.datos.gov.co/sites/default/files/CO_417_MANUAL_ETL.pdf Manual Creación de un ETL.

[20] http://herramientas.datos.gov.co/sites/default/files/CO_417_MANUAL_FEDERAR.pdf Manual de Federación.

[21] <http://www.odata.org/>

Se puede acceder a los conjuntos de datos del portal, incluidos los conjuntos de datos privados, a través de un punto final único OData que permite a los usuarios conectarse sin problemas a sus datos a través de diferentes herramientas como: Power Query, Excel 2010 o 2013 y Power BI Desktop. [22]

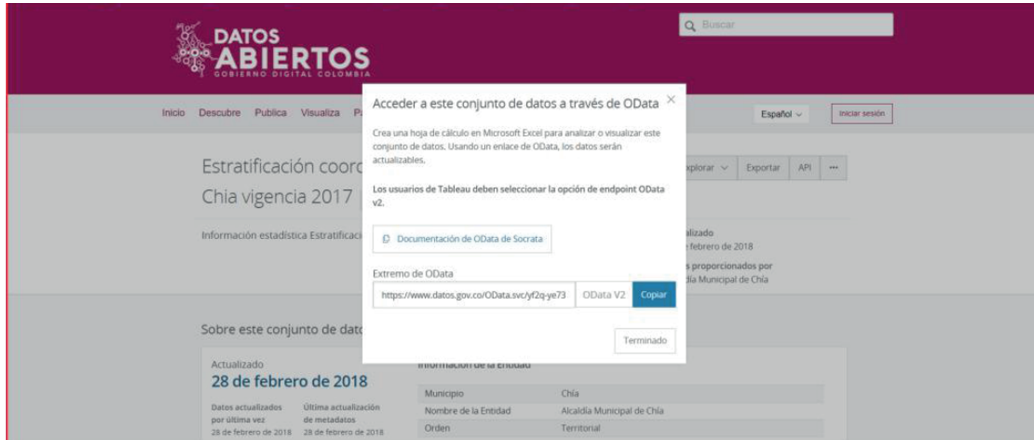


Imagen 25. Acceso al conjunto de datos a través de ODATA

SODA API: interfaz de programación de aplicaciones que permite su uso con diferentes lenguajes de programación. El portal de datos abiertos de Colombia, www.datos.gov.co cuenta con la plataforma SOCRATA, a través de la cual los desarrolladores de aplicaciones consumen los datos en línea directa y automáticamente desde sus desarrollos, sin necesidad de hacer el proceso de descarga y carga manualmente.



Imagen 26. Acceso al conjunto de datos a través de SODA API

[22] <https://support.socrata.com/hc/en-us/articles/115005364207-Access-Socrata-Data-using-Odata>

Ejemplo: aplicaciones de inteligencia de negocios, aplicaciones de visualización de datos, aplicaciones geográficas e incluso Excel.

En el siguiente enlace encontrará las instrucciones para acceder a los datos publicados mediante diferentes aplicaciones que soportan el uso de OData:

<https://support.socrata.com/hc/en-us/sections/360000042288-Access-Socrata-Datasets-using-Third-Party-Tools>

Para ampliar la información sobre la API del portal de datos abiertos de Colombia, consulte el siguiente enlace:

<http://herramientas.datos.gov.co/es/content/desarrollar-usando-los-datos>

SELLO DE EXCELENCIA



Para crear los más altos estándares de calidad, se plantea y se desarrolla el modelo del **Sello de Excelencia del Gobierno Digital** [23], el cual fue creado a través del decreto 2573 del año 2014 y adoptado en la resolución 2405 de 2016, que lo define como un modelo de certificación que busca garantizar la alta calidad de los servicios digitales del Estado colombiano.

El Sello de Excelencia está organizado en 3 categorías:

- **Categoría servicios en línea:** define los requisitos de calidad de los trámites y servicios en línea en cuanto a disponibilidad, seguridad, soporte, acceso, usabilidad, multicanalidad e interoperabilidad.
- **Categoría capacidades de gestión de TI:** establece las capacidades institucionales y operativas de las entidades públicas que tienen como propósito asegurar el adecuado funcionamiento de sus recursos de TI.
- **Categoría gobierno abierto:** establece los requisitos de calidad de los conjuntos de datos abiertos y plataformas de participación que propicien un gobierno transparente, abierto y participativo.
- **Categoría de Territorios y Ciudades Inteligentes:** certifica las iniciativas cuyo propósito sea solventar una problemática de carácter urbano, social, ambiental, económica y política, a través de Tecnologías de Información.

Los Requisitos para ser certificado por el Sello de Excelencia Gobierno Digital de Colombia se describen a continuación:

[23] <http://www.sellodeexcelencia.gov.co/banner/2>

Requisitos para ser certificado por el Sello de Excelencia Gobierno Digital de Colombia	
Etapas	Requisitos
Nivel 1	
Preparación para publicar el Conjunto de datos	<ol style="list-style-type: none"> 1. Identificación y priorización del conjunto de datos para su publicación. 2. Conjunto de datos completo y disponible en lotes. 3. Conjunto de datos actualizado. 4. Conjunto de datos documentado a través de metadatos.
Acceso y uso del conjunto de datos	<ol style="list-style-type: none"> 5. Conjunto de datos publicado en el portal de datos abiertos del Estado colombiano. 6. Conjunto de datos disponible para cualquier usuario de forma libre y sin restricciones. 7. Acceso al conjunto de datos de manera gratuita por parte del usuario. 8. Acceso al conjunto de datos sin ninguna restricción técnica y legal por parte del usuario. 9. Descarga del conjunto de datos en formatos libres por parte del usuario.
Evaluación y mejora	<ol style="list-style-type: none"> 10. Mecanismos para que los usuarios del conjunto de datos reporten hallazgos, comentarios o sugerencias.
Nivel 2	
Preparación para publicar el conjunto de datos	<ol style="list-style-type: none"> 1. El conjunto de datos cumple con los principios de legalidad para su publicación. 2. Análisis e identificación de la entidad de la vinculación de sus datos con otro(s) conjunto(s) de datos. 3. Cumplimiento de los criterios de calidad de los estándares nacionales e internacionales por parte del conjunto de datos geográficos.
Acceso y uso del conjunto de datos	<ol style="list-style-type: none"> 4. Divulgación y promoción por parte de la entidad responsable del conjunto de datos publicado para impulsar su uso. 5. La entidad responsable permite que terceros busquen automaticamente, recuperen o consuman datos, habilitando mecanismos de acceso a los datos por medio de Interfaces de Programación de Aplicaciones (APIs).
Evaluación y mejora	<ol style="list-style-type: none"> 6. La entidad desarrolla una estrategia que permite mejorar la calidad de los datos publicados a partir de la realimentación de sus usuarios. 7. Evaluación de los datos publicados y de la satisfacción de los usuarios por parte de la entidad.
Nivel 3	
Preparación para publicar el conjunto de datos	Todos los requisitos establecidos para el Nivel 1 y Nivel 2
Acceso y uso del conjunto de datos	<ol style="list-style-type: none"> 1. La entidad cuenta con mecanismos de publicación automática de los datos (ETL-proceso de extracción, transformación y carga o scripts) desde sus fuentes principales de información. 2. En caso de que aplique, el conjunto de datos debe de estar enlazado o vinculado con otro(s) conjunto(s) de datos.

Imagen 27. Requisitos del Sello de excelencia Gobierno Digital Colombia [24]

[24] http://estrategia.gobiernoenlinea.gov.co/623/articles-17464_recurso_14.pdf



El Sello de Excelencia certifica productos y servicios de entidades públicas que han liberado sus datos en el portal de datos abiertos de Colombia.

www.datos.gov.co.

Dicha certificación ha sido otorgada a diferentes conjuntos de datos publicados por entidades del orden nacional y territorial.^[25]

Los instrumentos y herramientas que facilitan la implementación del Sello de Excelencia los encuentra en su plataforma web:

<http://www.sellodeexcelencia.gov.co/>.

[25] <http://www.sellodeexcelencia.gov.co/certificados>



GLOSARIO

Anonimización:

es un proceso aplicado a un conjunto de datos con el fin de impedir la identificación, individualización y caracterización de los individuos sujetos en la producción de los datos. Esta se realiza por razones de seguridad para evitar posibles fines maliciosos que se le pueda dar a los datos publicados.

API: una Interfaz de Programación de Aplicaciones (Application Programming Interface) es un conjunto de protocolos, servicios y métodos de comunicación entre varios componentes de software. En las plataformas de datos abiertos, las API's proveen un servicio de consumo de datos que puede ser accedido por cualquier desarrollador que requiera dichos datos para alimentar una aplicación o

sistema informático. Las colecciones pueden tener un **enfoque temático u orgánico**, dependiendo de si están organizadas por un tema específico (clima, educación, transporte, etc.) o si están organizadas por la entidad que los publica, como por ejemplo "División de Transporte" de un municipio.[27]

Catálogo de datos:

es la parte central de un portal de datos abiertos y contiene un listado de todas las tablas publicadas con una descripción del contenido de la base de datos, el nombre de la agencia responsable, la frecuencia de actualización, el número de veces que se ha visitado, la información técnica para conectarla con aplicaciones informáticas y un espacio para los comentarios de los usuarios.

Algunos portales incorporan también un área de valoración de la calidad de la base de datos.[26]

Conjunto de datos -Dataset-:

es una colección de registros discretos representados en estructuras de datos generalmente tabulares (filas y columnas) que pueden ser accedidos y utilizados individualmente o en combinación.

Datos abiertos -Open Data-:

son datos digitales que son puestos a disposición con las características técnicas y jurídicas necesarias para que puedan ser usados, reutilizados y redistribuidos libremente por cualquier persona, en cualquier momento y en cualquier lugar.

[26] <https://blogs.iadb.org/abierto-al-publico/2014/11/20/que-es-un-portal-de-datos-abiertos-y-para-que-sirve/>

Distribuciones de los datos:

representa una forma accesible de datos en un catálogo de datos: un archivo descargable, una fuente RSS o un servicio web que proporciona los datos.

Interoperabilidad:

es la “capacidad de dos o más sistemas o componentes para intercambiar información y utilizar la información intercambiada”: ISO 25010.

Lenguaje común de**Intercambio de Información:**

es el estándar definido por el Estado colombiano para el intercambio de información entre sus dependencias y entidades.

Metadatos-Metadata:

son información adicional que describe, explica y facilita datos importantes acerca del conjunto de datos.

Estos permiten el acceso a los recursos de información y la reutilización de la misma. Los metadatos son fundamentales para la interoperabilidad de la información y para el cruce de datos en las instituciones públicas y privadas.

La estandarización de los campos de metadatos que se deben diligenciar en un dataset, posibilita a los usuarios comprender mejor el conjunto de datos y hacer mucho más sencillo ese cruce de información entre las entidades.

Visualizaciones de datos:

la sección de visualizaciones es un área donde se pueden publicar **gráficos o vistas de datos** que hayan creado los administradores o los usuarios, si la aplicación lo permite. Algunas plataformas vienen con módulos integrados para generar visualizaciones básicas y, en la mayoría de los casos, estas plataformas se conectan con otras herramientas externas que permiten hacer visualizaciones más vistosas y complejas. [28]

[27] <https://blogs.iadb.org/abierto-al-publico/2014/11/20/que-es-un-portal-de-datos-abiertos-y-para-que-sirve/>

[28] <https://blogs.iadb.org/abierto-al-publico/2014/11/20/que-es-un-portal-de-datos-abiertos-y-para-que-sirve/>