



**El futuro digital
es de todos**

**Gobierno
de Colombia
MinTIC**

**Modelo de Uso:
Calidad de Datos.**



Ministerio de Tecnologías de la Información y las Comunicaciones
Viceministerio de Economía Digital
Dirección de Gobierno Digital

Equipo de trabajo – Gobierno Digital
Luisa Fernanda Medina
Adriana Vargas Tamayo
Libertad Córdoba Varón
Benjamín Macías Argote
Ramiro Alberto Chaparro Vargas
Carlos Julio León Caicedo

Desarrollo técnico del documento
Harold Martínez – Grupo Kriterion
Carolina Gómez – Grupo Kriterion

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

Versión	Observaciones
Versión 1 octubre 2019	Planteamiento metodológico y conceptual de la estrategia de Calidad de Datos

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

Metodología y modelo de uso de Datos Abiertos



Esta Guía para el uso y aprovechamiento de Datos Abiertos en Colombia de la Dirección de Gobierno Digital se encuentra bajo una [Licencia Creative Commons Atribución 4.0 Internacional](https://creativecommons.org/licenses/by/4.0/).



Tabla de Contenido

1.Introducción.....	4
2.Descripción del caso de uso.....	5
3.Criterios de Calidad e Interoperabilidad.....	6
4.Automatización y Publicación.....	8
5.Operatividad del Modelo.....	9
5.1. Identificando los Conjuntos de Datos.....	9
5.2. Obteniendo la información de los conjuntos de datos y su metadata.....	9
5.3. Análisis de Calidad de Datos.	10
5.3.1. Indicador de Actualidad.....	10
5.3.2. Indicador de Trazabilidad.....	11
5.3.3. Indicador Conformidad.....	12
5.3.4. Indicador de Exactitud.....	13
5.3.5. Indicador de Completitud.....	14
5.3.6. Indicador de Consistencia.....	15
5.3.7. Indicador de Portabilidad.....	16
5.3.8. Indicador de Credibilidad.....	17
5.3.9. Indicador de Comprensibilidad.....	18
5.3.10. Indicador de Disponibilidad.....	18
5.4. Restricciones de Medición.....	19
6.Referencias.....	21

Índice de Tablas

Tabla 1-Criterios de calidad.....	6
Tabla 2-Reglas para el cálculo de indicadores de Actualidad.....	10
Tabla 3-Reglas para el cálculo de indicadores de trazabilidad.....	11
Tabla 4-Reglas para el cálculo de indicadores de conformidad.....	12
Tabla 5-Reglas para el cálculo de indicadores de exactitud.....	13
Tabla 6-Reglas para el cálculo de indicadores de completitud.....	14
Tabla 7-Reglas para el cálculo de indicadores de consistencia.....	15
Tabla 8-Reglas para el cálculo de indicadores de portabilidad.....	16
Tabla 9-Reglas para el cálculo de indicadores de credibilidad.....	17
Tabla 10-Reglas para el cálculo de indicadores de comprensibilidad.....	18
Tabla 11-Reglas para el cálculo de indicadores de disponibilidad.....	19
Tabla 12-Códigos de respuesta más comunes.....	19
Tabla 13-Indicadores de restricciones de medición.....	20

Índice de Ilustraciones

Ilustración 1-Arquitectura general de automatización.....	8
-----------------------------------------------------------	---



1. Introducción

La Estrategia de Gobierno en línea está liderada por el MinTIC, en coordinación con el DAFP (Departamento Administrativo de la Función Pública), y el DNP (Departamento Nacional de Planeación), y pretende el uso eficiente de las TIC para el cumplimiento de los objetivos del Gobierno Nacional. Según se ha publicado en el artículo Open Data Readiness Assessment, los datos abiertos son un componente fundamental en la estrategia del Gobierno en línea, y como tal se deben establecer los elementos habilitadores necesarios para lograr el objetivo de disponer de datos en línea para su reutilización: identificación de orígenes de datos, caracterización de usuarios de información, interoperabilidad, diálogo permanente y fomento de la reutilización (Colombia, Gobierno de, 2015).

De acuerdo con la Guía de Estándares de Calidad e Interoperabilidad de los Datos Abiertos del Gobierno de Colombia (MinTIC), además de lograr que las instituciones publiquen los datos producto del resultado de su trabajo, es necesario que esto se haga bajo estándares que garanticen la calidad y permitan la interoperabilidad de los datos.

El siguiente documento busca describir y explicar el proceso de evaluación de la calidad de la información disponible relacionada con la contratación pública en Colombia, en el marco del proyecto de Transparencia en la contratación pública, liderado por el Gobierno nacional.



2. Descripción del caso de uso

En el año 2018, se publicó por parte de la *Word Wide Web Foundation*¹, el informe del barómetro de los datos abiertos, en el que se presentan los resultados de la evaluación de 30 gobiernos que han asumido compromisos específicos para promover estos datos, ya sea mediante la adopción de la Carta de los Datos Abiertos o, como miembros del G20, mediante su adhesión a los Principios de los Datos Abiertos para la lucha contra la Corrupción del G20 (Foundation, World Wide Web, 2018).

Como parte de las conclusiones obtenidas, el informe plantea una serie de recomendaciones específicas que los gobiernos pueden llevar a cabo en áreas clave, entre las que se encuentra la posibilidad de construir y consolidar infraestructuras para los datos abiertos, mejorando la calidad y la interoperabilidad de los datos mediante prácticas eficaces de manejo de los datos y sistemas de gestión adecuados, para administrar datos abiertos, invirtiendo así en la capacitación y en la adquisición de competencias en materia de datos.

Al respecto, señala el informe que los gobiernos deberían tratar de publicar conjuntos de datos de calidad de manera uniforme en todos los organismos y departamentos, para que sea más fácil comprender y utilizar esos datos. No obstante, casi la mitad de los gobiernos líderes evaluados carecen aún de directrices, estándares técnicos o procedimientos de gestión integrales.

En tal sentido, el Gobierno Nacional ha preparado la Guía de Estándares de Calidad e Interoperabilidad de los Datos Abiertos del Gobierno de Colombia (MinTIC), con el objetivo de 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*.

Con base en las definiciones y prácticas establecidas en esta guía, se llevará a cabo el análisis de la información de contratación pública a fin de establecer el nivel de cumplimiento de los criterios de calidad.

¹ Organización para promover la web abierta como un bien público y un derecho básico



3. Criterios de Calidad e Interoperabilidad

Según se define en la Guía de Estándares de Calidad e Interoperabilidad de los Datos Abiertos del Gobierno de Colombia, los datos abiertos de alta calidad son una condición previa para analizar, reutilizar y garantizar el valor de los datos. Por lo anterior, es importante evaluar la calidad de los datos para verificar su exactitud, confiabilidad y aptitud.

Para lo anterior se han definido 16 criterios de calidad seleccionados a manera de estándar. Con la aplicación de éstos, 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:

Tabla 1-Criterios de calidad

Criterio	Definición Guía Interoperabilidad
Confidencialidad	Los datos solo deben ser accedidos por personas autorizadas para proteger adecuadamente la información reservada y clasificada.
Relevancia	Los datos publicados deben ser de utilidad para los usuarios. En este concepto, los datos tienen más o menos relevancia de acuerdo con el caso que se esté analizando.
Actualidad	Vigencia y actualidad de los datos publicados.
Trazabilidad	Histórico del conjunto de datos disponible: fechas de creación, publicación y actualizaciones.
Conformidad	Cumplimiento de lineamientos y estándares vigentes como DCAT (Data Catalog), para la descripción de la Metada.
Exactitud	Datos diligenciados correctamente.
Compleitud	Información completa
Consistencia	Datos coherentes y libres de contradicción.
Precisión	Nivel de desagregación en que están publicados los datos, con respecto al nivel de desagregación en que fueron generados.
Portabilidad	Formatos sin restricciones para la reutilización de los datos. Formatos no recomendados son PDF, DOC
Credibilidad	Información veraz y confiable para los usuarios.
Comprensibilidad	Características que permiten al usuario leer e interpretar los datos.
Accesibilidad	Herramientas tecnológicas que garanticen el acceso al usuario que lo requiera.
Eficiencia	Capacidad de la plataforma de realizar análisis y descargas de los datos, con unos niveles de desempeño y tiempos esperados.
Recuperabilidad	Atributos que permiten mantener y preservar un nivel específico de operaciones y de calidad.
Disponibilidad	Garantiza que los usuarios autorizados tengan acceso a la información y a otros activos de información asociados en el lugar,



Criterio	Definición Guía Interoperabilidad
	momento y forma en que es requerido. ¿Se puede acceder a los conjuntos de datos del portal?

Fuente: Elaboración Propia

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 (MinTIC).

4. Automatización y Publicación

El portal de Datos Abiertos permite la publicación de conjuntos de datos públicos, los cuales pueden ser automatizados a través de procesos de Extracción, Transformación y Cargue (ETL). A continuación, se presenta la arquitectura a seguir para su cargue y automatización.

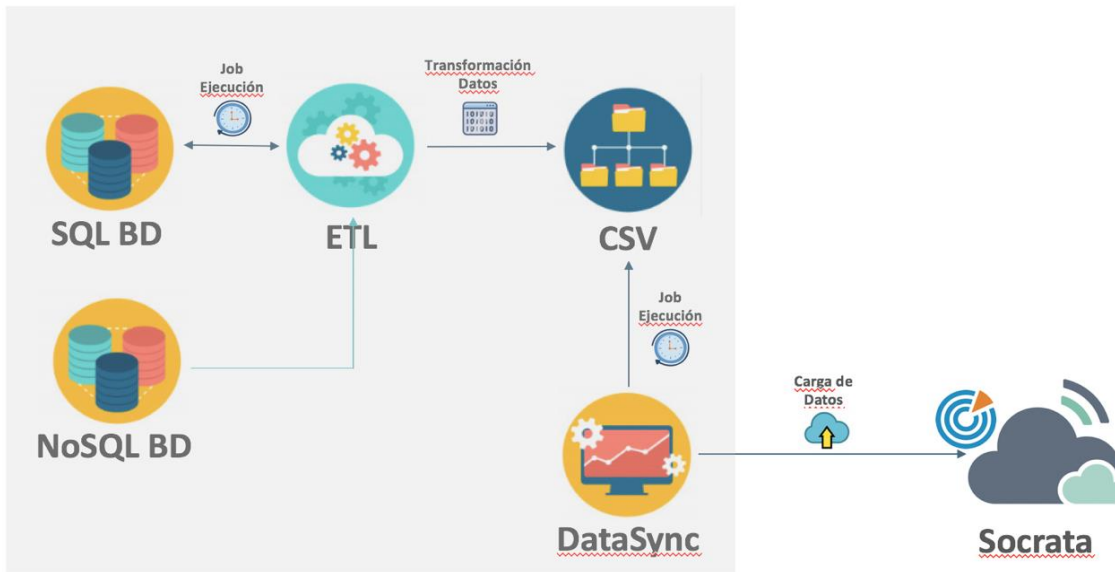


Ilustración 1-Arquitectura general de automatización

En el anterior diagrama se ven los siguientes componentes:

1. Los orígenes de la información pueden estar en las bases de datos estructuradas o *SQL*, también en bases de datos no estructuradas o *NoSQL*.
2. Para que los orígenes de datos puedan ser cargados en el portal de datos abiertos (Socrata), se deben generar en formato *CSV*, es decir, en un archivo plano separado por comas o punto y coma. Para esto se deben generar procesos de extracción de la información, transformación en un archivo plano y finalmente cargarlo en un directorio.
3. Una vez el conjunto de datos esté en formato *CSV*, se automatiza el cargue a través de la utilización de la herramienta *DataSync*.

El detalle del proceso de automatización se puede encontrar en la Guía de Automatización Versión 1.0. Donde se explica detalladamente los pasos necesarios para lograr publicar un conjunto de datos en el portal de datos.gov.co



5. Operatividad del Modelo

El alcance de evaluación del Modelo de Calidad de Datos está asociado a los conjuntos de datos presentes en el portal de Datos Abiertos del Estado colombiano, e identificados como públicos.

5.1. Identificando los Conjuntos de Datos

A continuación, se describen los pasos para lograrlo:

1. Hacer una consulta a <http://api.us.socrata.com/api/catalog/v1/domain> con la información de todos los dominios que se encuentran en esta plataforma, en donde se indica el nombre del dominio y la cantidad de datos que cada uno tiene.
2. Con el listado anterior, se debe buscar el dominio www.datos.gov.co para identificar la cantidad actual de datos que tiene el portal del estado colombiano.
3. Una vez identificado el número de conjuntos de datos, se debe buscar los IDs para cada uno de estos datos, haciendo una búsqueda con el siguiente link: <https://api.us.socrata.com/api/catalog/v1?domains=datos.gov.co>
4. La búsqueda anterior trae la información en formato *JSON* de los conjuntos de datos, en donde se puede extraer el “*id*” de cada uno ellos. Este “*id*” será utilizado para luego descargar la información del conjunto de datos y su metadata, a través de un llamado *HTTPS GET*. El “*id*” o identificador contiene 8 caracteres separados, en dos conjuntos de 4 caracteres por un guion: (ejemplo: abcd-1234)

5.2. Obteniendo la información de los conjuntos de datos y su metadata.

Para poder hacer la evaluación de la calidad de datos, se requieren la información del conjunto de datos y además de la metadata asociada a este conjunto de datos. Los pasos para conseguirlo se describen a continuación:

1. El usuario debe estar registrado en el portal de datos abiertos y tener un usuario, contraseña y, además, un *token* de acceso. Si el usuario no utiliza el token entregado por la misma plataforma, el número de registros a descargar por cada conjunto de datos estará limitado a dos mil (2.000)
2. Con el usuario, contraseña y *token* de acceso se podrá descargar la información necesaria para hacer la evaluación de calidad. Para lograrlo, se pueden utilizar diferentes lenguajes de programación tales como, *Python*, *Java*, *JavaScript*, *R*, *Ruby*,



entre otros, para así poder descargar la información. En nuestro caso, hemos utilizado *Python* a través de la librería *Sodapy*.

3. Para descargar la información del conjunto de datos, utilizamos la instrucción: *client.get ("id")* con el id del conjunto específico de datos que queramos descargar, si se requiere limitar los registros a descargar, se puede utilizar adicionalmente el parámetro "*limit*" de la siguiente manera: *client.get ("id", limit=5000)*, donde se limitaría a 5.000 registros en la descarga.
4. Para descargar la información de la metadata del conjunto de datos específico, se debe utilizar la instrucción *client.get_metadata("id")*, siendo "*id*" el identificador del conjunto de datos. Esta información será la materia prima para realizar los análisis de calidad de datos.
5. El lenguaje utilizado para el análisis de calidad de datos fue Python versión 3.6. En el siguiente repositorio se encuentra el código utilizado con este propósito: <https://github.com/softwarepublicomintic/calidaddatos.git>

5.3. Análisis de Calidad de Datos.

A partir de los conjuntos de datos descargados anteriormente y su metadata respectiva, se realizará la evaluación de calidad. A continuación, se definen cada uno de los atributos y su modelo de evaluación.

5.3.1. Indicador de Actualidad

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

Tabla 2-Reglas para el cálculo de indicadores de Actualidad

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

Fuente: Elaboración Propia



- **Cálculo de indicador**

Paso 1: identificar la frecuencia de la actualización del conjunto de datos, a partir de la información en la metadata. La fecha de actualización se encuentra en la ruta *JSON: custom_fields > Información de Datos > Frecuencia de Actualización*.

Paso 2: identificar cuál fue la última fecha de actualización de los datos, la cual se encuentra en el siguiente campo *JSON: rowsUpdatedAt*.

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

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

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

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

5.3.2. Indicador de Trazabilidad

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

Tabla 3-Reglas para el cálculo de indicadores de trazabilidad



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

Fuente: Elaboración Propia

- **Cálculo de indicador**

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

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

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

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

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

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

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

5.3.3. Indicador Conformidad

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

Tabla 4-Reglas para el cálculo de indicadores de conformidad

Criterio	Definición Guía Interoperabilidad	Medible	Métrica por Conjunto de Datos
Conformidad	Cumplimiento de lineamientos y estándares vigentes: 1) Metadatos 2) Metadatos: que expliquen el conjunto de datos y cumplan <i>DCAT</i>	Si	Porcentaje de completitud de metadata <i>DCAT (Data Catalog)</i> del conjunto de datos: nombre, descripción, categoría, fecha de actualización, correo electrónico, entidad



Criterio	Definición Guía Interoperabilidad	Medible	Métrica por Conjunto de Datos
			publicadora e identificador del conjunto de datos.

Fuente: Elaboración Propia

- **Cálculo de indicador**

Paso 1: se obtienen los campos de la metadata asociado al nombre (“name”), descripción (“description”), categoría (“category”), fecha de actualización (“rowsUpdatedAt”), correo electrónico (“email”), entidad publicadora (“attribution”), identificador (“id”).

Paso 2: identificar si todos los campos están incluidos o no.

Paso 3: cálculo del indicador de acuerdo a la completitud de los datos: total de datos completos, dividido en total de campos (7), multiplicado por 10

5.3.4. Indicador de Exactitud

- **Reglas de Negocio:** para realizar el cálculo de los indicadores se definieron las siguientes reglas:

Tabla 5-Reglas para el cálculo de indicadores de exactitud

Criterio	Definición Guía Interoperabilidad	Medible	Métrica por Conjunto de Datos
Exactitud	Datos diligenciados correctamente.	Si	<p>1) Porcentaje de cumplimiento de tipo de dato, de acuerdo con el tipo de dato definido para la variable. Se promedia el resultado de la evaluación hecha a cada columna del conjunto de datos: cantidad de tipo de datos del tipo definido/ cantidad total de registros</p> <p>2) Porcentaje de datos diligenciados correctamente en el conjunto de datos. Se promedia el resultado de la evaluación hecha a cada columna del conjunto de</p>



Criterio	Definición Guía Interoperabilidad	Medible	Métrica por Conjunto de Datos
			datos: cantidad de datos correctos/ cantidad total de registros. Validación versus valores estándar de acuerdo con el campo establecido.

Fuente: Elaboración Propia

- **Cálculo del Indicador**

Paso 1: identificar los tipos de datos para cada una de las columnas del conjunto de datos, de acuerdo a la información de la Metadata.

Paso 2: a partir de la anterior identificación, ir a cada columna y medir a través de expresiones regulares, si el tipo de dato corresponde a lo especificado en la metada.

Paso 3: identificar la cantidad de datos que no corresponde al tipo definido en la metada.

Paso 4: identificar la cantidad de registros totales.

Paso 5: aplicar la siguiente fórmula:

$$\text{Exactitud} = (1 - (\text{datos que no coinciden} / \text{registros totales})) * 10$$

5.3.5. Indicador de Completitud

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

Tabla 6-Reglas para el cálculo de indicadores de completitud

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



Fuente: Elaboración Propia

• **Cálculo de indicador**

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

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

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

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

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

$$(1 - (\text{suma registros faltantes} / \text{suma registros totales})) \times 10$$

5.3.6. Indicador de Consistencia

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

Tabla 7-Reglas para el cálculo de indicadores de consistencia

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

Fuente: Elaboración Propia



• **Cálculo de indicador**

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

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

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

$$Columnas = (1 - (total\ registros\ repetidos / total\ registros)) \times 10$$

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

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

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

$$Filas = (1 - (total\ registros\ repetidos / numero\ registros\ totales)) \times 10$$

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

$$Indicador\ Consistencia = (Columnas + Filas) / 2$$

5.3.7. Indicador de Portabilidad

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

Tabla 8-Reglas para el cálculo de indicadores de portabilidad

criterio	Definición Guía Interoperabilidad	Medible	Métrica por Conjunto de Datos
Portabilidad	Formatos sin restricciones para la reutilización de los datos. Formatos no recomendados son PDF, DOC	Si	1) Cumple/ No Cumple (DOC o PDF) 2) Especificar formato del conjunto de datos y modo de acceso

Fuente: Elaboración Propia



Aclaración: El portal de datos abiertos www.datos.gov.co, permite la descarga de conjuntos de datos en CSV, TSV, RDF, XML, y RDS. Al acceder a través de la API a los conjuntos de datos, estos están en formato JSON.

- **Cálculo de indicador**

Paso 1: acceder al conjunto de datos a través de API. Comprobar disponibilidad.

Paso 2: convertir formato JSON a format String, a través de operación en Python `json.dumps()`

Paso 3 (Salida): si la conversión fue satisfactoria, entonces el indicador es 10, si no es posible hacer la operación, indica que existe un problema con el formato JSON en la descarga.

5.3.8. Indicador de Credibilidad

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

Tabla 9-Reglas para el cálculo de indicadores de credibilidad

Criterio	Definición Guía Interoperabilidad	Medible	Métrica por Conjunto de Datos
Credibilidad	Información veraz y confiable para los usuarios.	Si	Cumple/No cumple Cumple si tiene institución y usuario de contacto

Fuente: Elaboración Propia

- **Cálculo de indicador**

Paso 1: extraer la información de la institución, a través de la siguiente ruta en el formato JSON de la metadata: `permmission > users > displayName`

Paso 2: extraer la información del correo electrónico, a través de la siguiente ruta en el formato JSON de la metadata: `permmission > users > email`

Pase 3 (Salida): si cuenta con los datos de institución y correo electrónico, entonces el indicador será 10. Si alguno de los dos campos se encuentra vacío, entonces el indicador será 0.



5.3.9. Indicador de Comprensibilidad

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

Tabla 10-Reglas para el cálculo de indicadores de comprensibilidad

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

Fuente: Elaboración Propia

- **Cálculo de indicador**

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

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

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

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

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

$$\text{Comprensibilidad} = (1 - (\text{nombre coincidentes}/\text{columnas únicas})) \times 10$$

5.3.10. Indicador de Disponibilidad



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

Tabla 11-Reglas para el cálculo de indicadores de disponibilidad

Criterio	Definición Guía Interoperabilidad	Medible	Métrica por Conjunto de Datos
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.	Si	Cumple/No Cumple Puede accederse al conjunto de datos.

Fuente: Elaboración Propia

Aclaración: como el acceso a los conjuntos de datos y su metadata es través del API (HTTP GET), se pueden obtener diferentes códigos de respuesta. Los siguientes son los códigos de respuestas más comunes:

Tabla 12-Códigos de respuesta más comunes

Códigos de Estado de Respuesta HTTP	Descripción
200	Respuesta exitosa
204	Respuesta exitosa, pero sin contenido
400	El servidor no pudo interpretar la solicitud enviada.
401	Es necesaria la autenticación para acceder
403	No se tienen los permisos necesarios para acceder a la información.
405	El método de conexión se identificó, pero no está permitido.

Fuente: Elaboración Propia

- **Cálculo de indicador**

Paso 1: hacer llamado HTTP al conjunto de datos, a través del ID.

URL: <https://www.datos.gov.co/resource/idConjuntoDatos.json>

Paso 2 (Salida): si el estado de la respuesta es 200, entonces el indicador es 10, de lo contrario el indicador es 0.

5.4. Restricciones de Medición.



De acuerdo a lo identificado inicialmente, los siguientes atributos de calidad no pueden ser medidos por las restricciones o deficiones del portal de datos abiertos

Tabla 13-Indicadores de restricciones de medición

Criterio	Definición Guía Interoperabilidad	Medible	Métrica por Conjunto de Datos
Relevancia	Los datos publicados deben ser de utilidad para los usuarios.	Por defecto	No aplica, ya que depende del uso de los datos. No puede ser evaluado de manera aislada.
Precisión	Nivel de desagregación en que están publicados los datos, con respecto al nivel de desagregación en que fueron generados.	No	No es posible medir el grado de desagregación de un dato, sino a través del conocimiento de sus orígenes.
Accesibilidad	Herramientas tecnológicas que garanticen el acceso al usuario que lo requiera.	Por defecto	Se cumple por defecto. Inherentes al funcionamiento de la plataforma de datos abiertos
Eficiencia	Capacidad de la plataforma de realizar análisis y descargas de los datos, con unos niveles de desempeño y tiempos esperados.	No	Se cumple por defecto. Inherentes al funcionamiento de la plataforma de datos abiertos
Recuperabilidad	Atributos que permiten mantener y preservar un nivel específico de operaciones y de calidad.	No	No se cumple por defecto. La plataforma de datos.gov.co no permite ver conjuntos de datos previos, solamente el actual.
Confidencialidad	Los datos solo deben ser accedidos por personas autorizadas para proteger adecuadamente la información reservada y clasificada. Información debe ser anonimizada y generalizada	Por defecto	No aplica. Los datos analizados cumplen con la característica inicial de ser datos públicos.

Fuente: Elaboración Propia



6. Referencias

COLOMBIA, G. D. (2015). Open Data Readiness Assessment. Word Bank Group. Retrieved from https://estrategia.gobiernoenlinea.gov.co/623/articles-9407_evaluacion_apertura_datos.pdf

FOUNDATION, W. W. (2018). Webfoundation.gov. Retrieved from https://webfoundation.org/docs/2018/09/WF_ODB_Report2_Spanish_Screen.pdf

MinTIC, G. d. (s.f.). Guía de estándares de calidad e interoperabilidad de los Datos Abiertos del gobierno de Colombia. Retrieved from https://herramientas.datos.gov.co/sites/default/files/A_guia_de_estandares_final_0.pdf



**El futuro digital
es de todos**

**Gobierno
de Colombia
MinTIC**