Great Place To Work®
Certificado
OCT 2020-SEP 2021
COL

El futuro digital es de todos
MinTIC

CALIDAD DE DATOS ABIERTOS

Taller 2. Mejorando la calidad de datos

Septiembre 29 de 2021

Calidad de datos

AGENDA: Sesión de sensibilización

El futuro digital es de todos | MinTIC

Taller 2: Mejorando la Calidad del Dato

0 Introducción

1. Usuario publicador

    1.1. Calidad de metadatos

2. Tipos de errores comunes al publicar datos abiertos

3. Mejorando la calidad del dato, solución de errores comunes

4. Principios de la calidad para publicación de datos abiertos

5. Atributos de la calidad

6. Procesos en la calidad de los datos

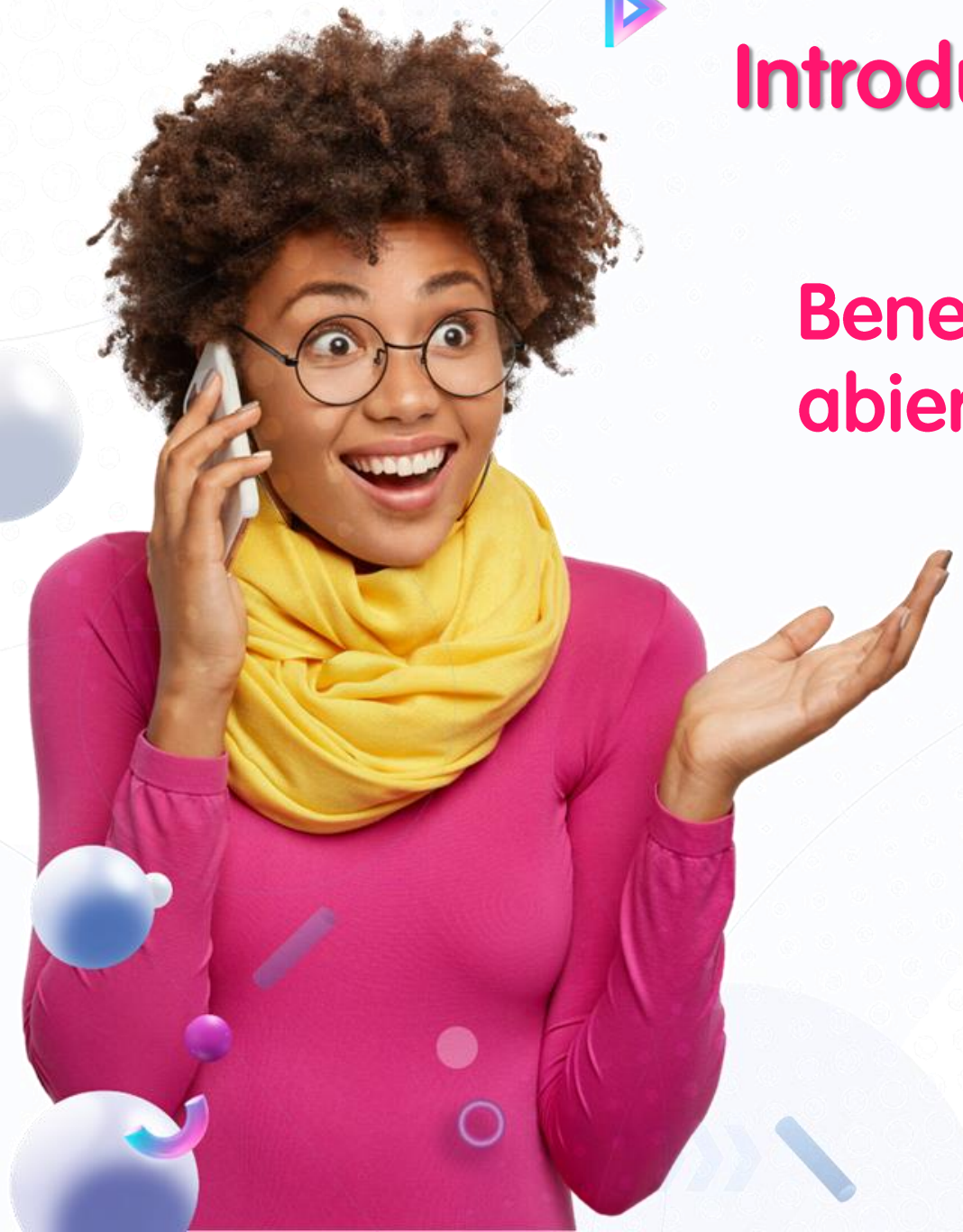
7. Planificación de la calidad del dato

8. Control y aseguramiento de la calidad

DATOS ABIERTOS
GOBIERNO DIGITAL COLOMBIA

Objetivos del taller

- Identificar los errores más comunes que se detectan en la publicación de los datos abiertos en el portal
- Facilitar la corrección de los errores detectados en la publicación
- Planificar la calidad del dato en cuanto a su validación, construcción y monitoreo de los datos empleados.
- Conocer los principios regulatorios en la calidad del dato



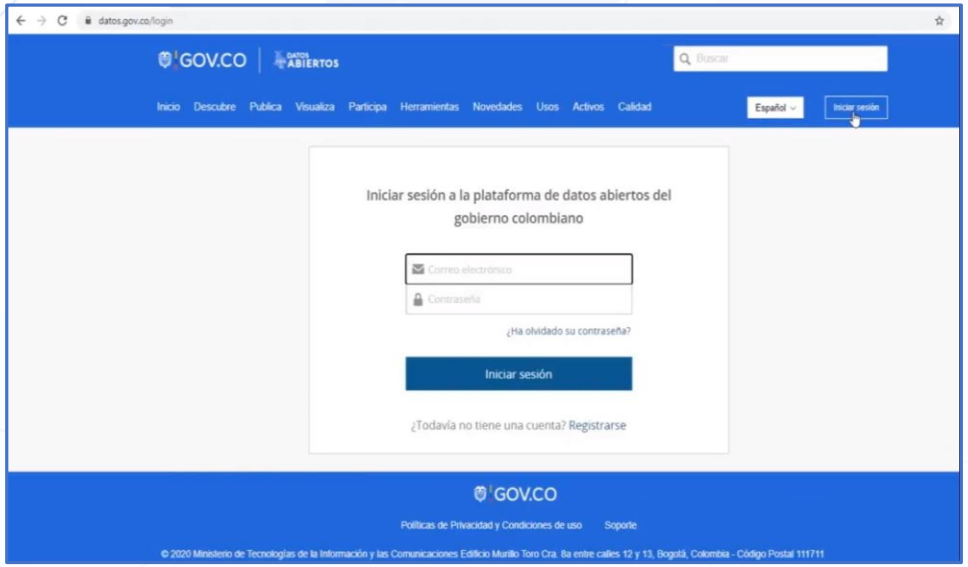
Introducción

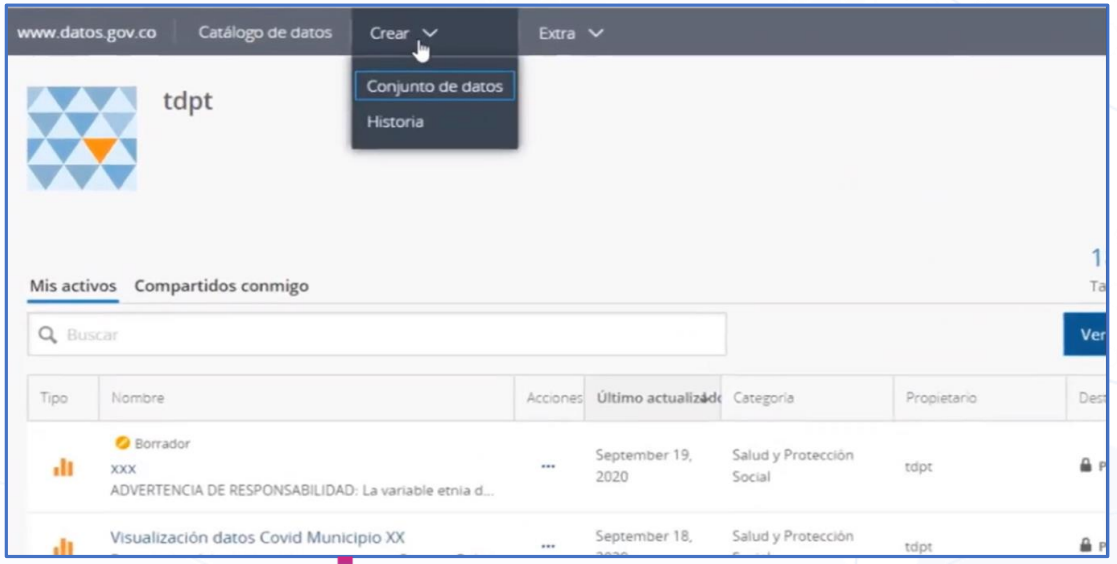
Beneficios de tener calidad en los datos abiertos

- ¿Qué significa calidad de datos?
- ¿Por qué debería aprender del tema y cómo incorporarlo en mi entidad?
- ¿Qué veremos en la siguiente sesión?

Usuario Publicador

Ingresando al Portal de datos abiertos



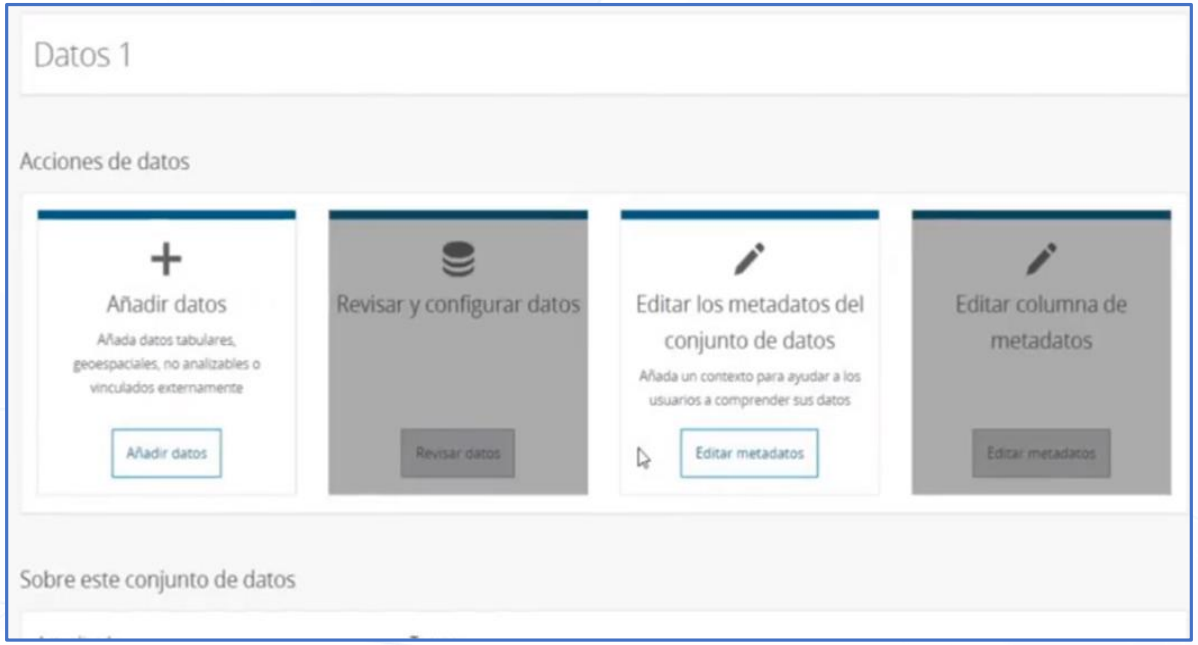


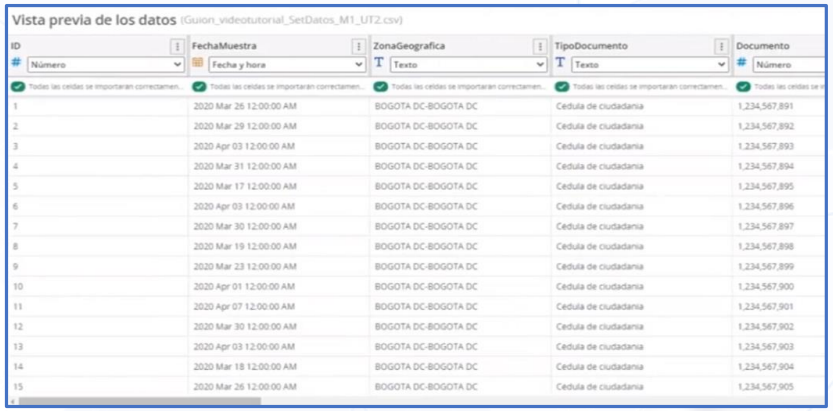
www.datos.gov.co

En el Portal

Usuario Publicador

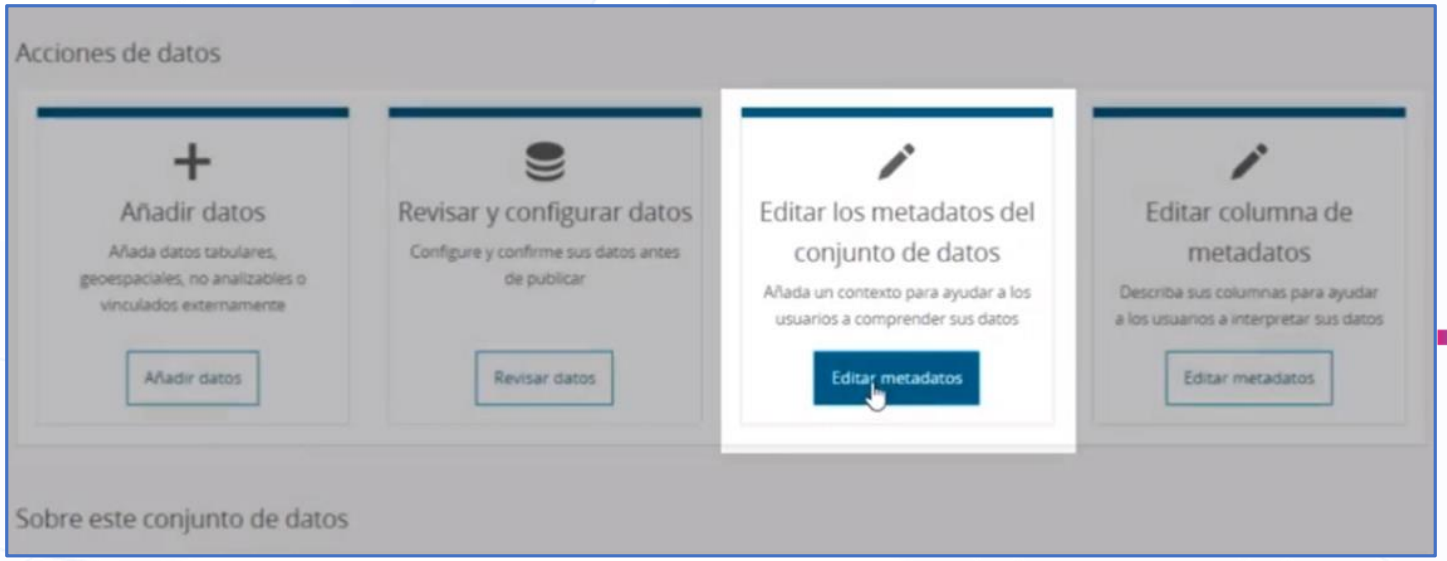
Ingresando al Portal de datos abiertos

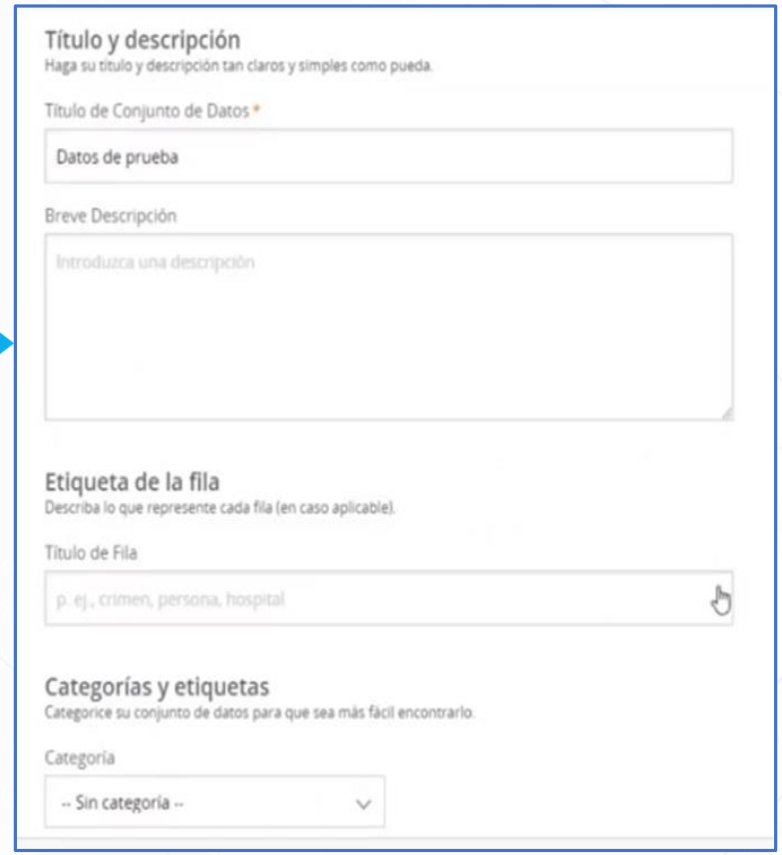




Calidad de los Metadatos

Ingresando al Portal de datos abiertos





Calidad de los Metadatos

Ingresando al Portal de datos abiertos

El futuro digital es de todos | MinTIC

Tipo	Nombre	Aprobación solicitada	Propietario
	Estados Financieros Grupo 1 Separados (2019) La Superintendencia de Sociedades pone a disposición...	August 24, 2021	amorales

Actualizado
24 de agosto de 2021

Última actualización de metadatos
24 de agosto de 2021

Fecha de cración
24 de agosto de 2021

Vistas
4

Descargas
0

Datos ofrecidos por
Superintendencia de Sociedades

Propietario de conjunto de datos
amorales

Contactar con dueño de conjunto de datos

Información de la Entidad

Área o dependencia	SUPERINTENDENCIA DE SOCIEDADES
Nombre de la Entidad	Superintendencia de Sociedades
Departamento	Bogotá D.C.
Municipio	Bogotá D.C.
Orden	Nacional
Sector	Comercio, Industria y Turismo

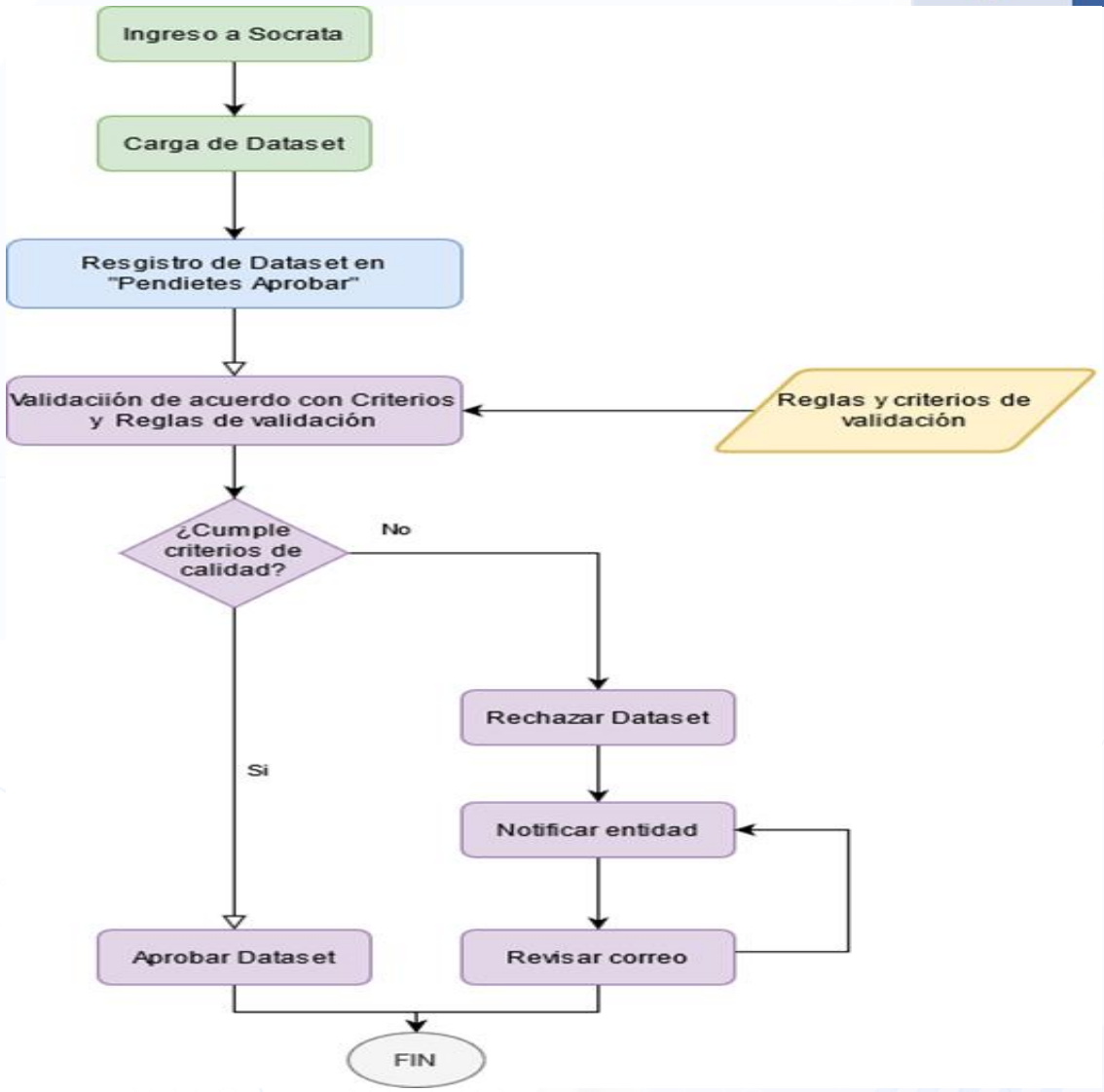
Información de Datos

Idioma	Español
Cobertura Geográfica	Nacional
Frecuencia de Actualización	Anual
Fecha Emisión (aaaa-mm-dd)	2021-08-24
URL Documentación	https://siis.ia.supersociedades.gov.co/#/
URL Normativa	http://www.supersociedades.gov.co/delegatura_aec/informes_empresariales/Paginas/sirfin.aspx

DATOS ABIERTOS
GOBIERNO DIGITAL COLOMBIA

Tipos de errores más comunes

- Flujo de Aprobación.



Tipos de errores más comunes al publicar Datos Abiertos

- Visualización del inventario de errores frecuentes, relacionados con los criterios de calidad dentro del portal de datos abiertos

En el Portal

www.datos.gov.co

Vista previa de la tabla

Explorar los datos | Crear visualización

Código de Error	Categoría de Er...	Tipo de Error	Descripción Error	Solución	Herramientas d...	Estado del Conj...
ERR001	Metadata Errada, Inc...	Titulo y Descripcion ...	El titulo y descripcio...	Se recomienda actua...	Ver video Capacitaci...	Si en el transcurso d...
ERR002	Metadata Errada, Inc...	La Metadata del conj...	La metadata esta inc...	Se recomienda actua...	Ver video Capacitaci...	Si en el transcurso d...
ERR003	Metadata Errada, Inc...	El campo de la meta...	El campo de la meta...	Se recomienda con e...	Ver video Capacitaci...	Si en el transcurso d...
ERR004	Error SIN Filas	Conjunto de datos n...	Actualmente su conj...	Se recomienda ELIMI...	Ver el video como p...	El administrador del ...
ERR005	Error Poca Filas	Conjunto de datos n...	El conjunto de datos ...	Dado a que el conju...	Guia de Datos Abiert...	El administrador del ...
ERR005_01	Error Poca Filas y Ag...	Conjunto de datos n...	El conjunto de datos ...	Se recomienda mejo...	Guia de Datos Abiert...	
ERR005_02	Error Poca Filas y Cla...	Conjunto de datos n...	El conjunto de datos ...	Se recomienda mejo...	Guia de Datos Abiert...	El administrador del ...
ERR005_02	Error Pocas Filas Turi...	Conjunto de datos n...	El conjunto de datos ...	Se recomienda para ...	Guia de Datos Abiert...	El administrador del ...
ERR007	Error Filas Vacias	Conjunto de datos c...	El conjunto de datos ...	Se recomienda hace...	Ver el video como p...	El administrador del ...
ERR008	Error Columnas	El conjunto de datos ...	Actualmente la estru...	Se recomienda elimi...	Ver el video como p...	El administrador del ...
ERR008_1	Error Poca Columnas	Conjunto de datos p...	El conjunto de datos ...	Se recomienda mejo...	Guia de Datos Abiert...	El administrador del ...
ERR008_2	Error Columnas mal ...	Conjunto de datos p...		Se recomienda hace...	Ver el video como p...	El administrador del ...
ERR009	Error Columnas	Error falta campo de...	El conjunto de datos ...	Se recomienda inclui...	Ver video Capacitaci...	El administrador del ...

< Anterior Siguiente > Mostrando calidad 1-13 de 38

https://www.datos.gov.co/Ciencia-Tecnolog-a-e-Innovaci-n/Errores-Frecuentes-de-Calidad-Conjuntos-de-datos-a/xbc7-65j4

Mejorando la calidad del dato, solución de errores comunes

El futuro digital es de todos | MinTIC

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

DATOS ABIERTOS
GOBIERNO DIGITAL COLOMBIA

Veamos el Excel con los errores detectados en el Portal de Datos Abiertos y sus Soluciones

Principios de la calidad

para publicar Datos abiertos

El futuro digital es de todos | MinTIC

Los datos son considerados abiertos si son cumplen con los siguientes principios de calidad:

- **Completos:** Todos los datos públicos deben estar disponibles. Los datos públicos no contemplan datos privados ni limitaciones de seguridad o privilegios.
- **Primarios:** Los datos deben ser recolectados en la fuente de origen, con el nivel de granularidad más alto posible, no en forma agregada ni modificada.
- **Oportunos:** Los datos deben estar disponibles tan rápido como sea necesario para garantizar el valor de los mismos.
- **Accesibles:** Los datos deben estar disponibles para el rango más amplio de usuarios y para el rango más amplio de propósitos.
- **Procesables por máquinas:** Los datos deben estar estructurados razonablemente para permitir un procesamiento automático.
- **No discriminatorios:** Los datos deben estar disponibles para cualquier persona, sin requerir un registro.
- **No propietarios:** Los datos deben estar disponibles en un formato sobre el cual ninguna entidad tiene un control exclusivo.
- **Libres de licencias:** Los datos no deben estar sujetos a ningún derecho de autor, patente, marca registrada o regulaciones de acuerdo de secreto. Se podrán permitir restricciones razonables de privacidad, seguridad o privilegios.

DATOS ABIERTOS
GOBIERNO DIGITAL COLOMBIA

Atributos de calidad

Perfilado y reporte alrededor de atributos de calidad del dato:

El Reporte de perfilado se utiliza para presentar avance en calidad contabilizando las instancias de atributo encontradas.

Para realizar el perfilamiento de datos se utiliza:

1. **Informe de perfilado**: este informe incluye perfilado de datos, de información y de la función pública (en el sector público).

    1.1. **En el perfilado de datos se obtienen de manera automática** las mediciones de datos nulos, duplicados, posibles campos en desuso, máscaras de formato del dato, validaciones de formato de datos, tipologías de casuísticas, etc. En este paso se detectan por ejemplo codificaciones incorrectas (por ejemplo en departamentos del Estado colombiano), tipologías no previstas (por ejemplo más de dos tipos de sexo: hombre, mujer, masculino, femenino, etc.).

    Este perfilado debe estar apoyado por una herramienta informática especializada para tal fin.

Atributos de calidad

1.2 **El perfilado de la información es un procedimiento semiautomático**. Este perfilado va un paso más allá que el perfilado de datos. Por ejemplo, si en el perfilado de datos se valida que una fecha sea correcta, en el perfilado de información se iría más allá, interpretando la fecha, tal vez como una fecha de nacimiento, y por lo tanto obteniendo nuevos análisis.

1.3 El perfilado de función pública permite implementar lógica compleja, por ejemplo, sobre resultados en campañas finalizadas en el sector público para stakeholders, o sobre el uso de reglas cruzadas para validación en cálculos de tarificación o facturación de servicios, etc.

Este perfilado de función pública se apoya tanto en los datos como en sus metadatos funcionales: interpretaciones en el sector público del dato, lógicas de validación y umbrales de calidad.

2. Informe final: este muestra la identificación de iniciativas preventivas para limpieza en los sistemas fuente y para detección de posibles inconsistencias en sus repositorios, así como la detección de iniciativas 33 Información Guía Técnica para validación cruzada durante la ejecución del perfilado de datos, con el fin de aplicarlas en la fase de diseño de la migración de datos.

Procesos

Calidad de Datos

El futuro digital es de todos

MinTIC

Los procesos de calidad de los datos está ligado a la limpieza de datos y siempre esta asociado a los objetivos determinados en el proceso de captura de datos.

Esto esta definido por las características del negocio de la organización: ¿para qué se están capturando datos? ¿qué tipo de datos se requieren? según las respuestas se requiere definir diversos procesos de limpieza o ninguno.

La limpieza de los datos incluye 5 actividades principales:

1. Depurar
2. Corregir
3. Estandarizar
4. Relacionar
5. Consolidar

DATOS ABIERTOS
GOBIERNO DIGITAL COLOMBIA

Planificación de la calidad

Fases

Planeación

- Reuniones de seguimiento a la calidad de los metadatos y datos del portal.
- Documentar el plan estratégico de calidad de metadatos y datos del portal.

Desarrollo

- Estimar ajustes y nuevos desarrollos solicitados a DataQ.
- Realizar pruebas unitarias y funcionales a DataQ.

Ejecución

- Ejecución de programa de automatización (mensual o quincenal).
- Comunicación con la entidad mediante correos (mensual o quincenal).
- Monitoreo del correo de datosabiertos@mintic.gov.co

Seguimiento

- Reunión mensual de seguimiento.
- Informe DataQ sobre el avance en la mejora de la calidad (mensual o quincenal).

El futuro digital es de todos
MinTIC

Objetivo

Establecer una estrategia en la cuál se defina la responsabilidad, la secuencia de ejecución e interacción de los procesos para la Gestión de la Calidad de los Metadatos y Datos del portal de datos abiertos del estado colombiano.

Planificación de la calidad

Ejecución de la calidad

El futuro digital es de todos

MinTIC

Semanal

El conjunto de datos contiene menos de cinco registros.

- El enlace del conjunto de datos externo no permite la descarga directa de un conjunto de datos en formatos válidos: csv, xls, xlsx, json, kml, kmz y zip (shapefile de ESRI)
- El enlace que se encuentra en el campo de la metadata URL documentación y/o URL normativa, direcciona a una página no disponible
- Metadata Incompleta y/o vacía
- Conjunto con una columna
- Carga de archivos no admitidos como Datos Abiertos

Quincenal

Ajustes automáticos municipio, departamento, entidad, formato de fechas.

- El conjunto de datos no se encuentra actualizado a la fecha, de acuerdo al campo de la Metadata frecuencia de actualización y la última fecha de actualización del conjunto de datos
- El enlace del conjunto de datos externo direcciona a una página no disponible
- Modificar el nombre del usuario
- Conjunto con columnas de solo texto
- Conjunto de datos no tiene filas con información

DATOS ABIERTOS
GOBIERNO DIGITAL COLOMBIA

Control y aseguramiento

De la calidad de Datos

Descubrimiento de datos: proceso de búsqueda, recopilación, organización y notificación de metadatos.

Perfilado de datos: proceso de analizar los datos en detalle, comparándolos con sus metadatos, calculando estadísticas de datos e informando de las medidas de calidad de los datos que se deben aplicar en cada momento.

Reglas de calidad de datos: se orientarán a optimizar el nivel de calidad de los activos informacionales de la organización y, para ello, se basarán en los requisitos de negocio aplicables, las reglas comerciales y técnicas a las que deben adherirse los datos.

Monitorización de la calidad de los datos: la mejora continua requiere de un esfuerzo de seguimiento, que permita comparar los logros con los umbrales de error definidos, la creación y almacenamiento de excepciones de calidad de datos y la generación de notificaciones asociadas.

Reporting de calidad de datos: está relacionado con los procedimientos y herramientas empleadas para informar, detallar excepciones y actualizar las medidas de calidad de datos en curso.

Limpieza de datos: se ocupa de la corrección en curso de las excepciones y problemas de calidad de datos según son notificadas.

El futuro digital es de todos

MinTIC

PREGUNTAS

DATOS ABIERTOS
GOBIERNO DIGITAL COLOMBIA

El futuro digital es de todos

MinTIC

¡MUCHAS GRACIAS!

ALIRIO HUMBERTO OTÁLORA VIVARES
Ingeniero de Sistemas con énfasis en Software
Científico de Datos

DATOS ABIERTOS
GOBIERNO DIGITAL COLOMBIA