

Great
Place
To
Work®

Certificado
OCT 2020-SEP 2021
COL



El futuro digital
es de todos

MinTIC

CALIDAD DE DATOS ABIERTOS

1. Calidad del Dato

Septiembre 22 de 2021



Calidad de datos

AGENDA: Sesión de sensibilización

Módulo 0: Introducción

Tema 1: Bienvenida.

- 0.1 Bienvenida
- 0.2 Objetivos
- 0.3 ¿qué son los Datos Abiertos?
- 0.4 Beneficios de tener alta calidad en datos abiertos
- 0.5 ¿Por que abrir Datos Abiertos con calidad?

Módulo 1: Calidad del Dato

- 1.1 ¿Qué es calidad del dato?
- 1.2 Características y Criterios de la calidad del dato.
 - 1.2.1 Exactitud, completitud, consistencia, Precisión, Duplicidad
 - 1.2.2 Confidencialidad, Relevancia, Actualidad, Trazabilidad, Conformidad
 - 1.2.3 Portabilidad, Credibilidad, Comprensibilidad, Accesibilidad
- 1.3 Plan de calidad del dato
- 1.4 Herramientas de la Calidad del dato
- 1.5 Usuario Publicador y Calidad de los Metadatos



El futuro digital
es de todos

MinTIC



Objetivos



- Conocer la iniciativa de datos abiertos del Gobierno de Colombia
- Identificar los estándares mínimos de calidad que deben tener los datos abiertos tanto en los metadatos como en el contenido de los conjuntos de datos
- Comprender el concepto de calidad del dato, sus características y aplicación de los criterios de calidad.
- Conocer los pasos para realizar el plan de calidad de datos.
- Identificar los criterios de calidad del dato para su validación, construcción y monitoreo de los datos empleados.



▶ ¿Qué son los datos abiertos?





▶ Beneficios de datos abiertos

- Desarrollo de capacidades para el uso estratégico de la información.
- Toma de decisiones basadas en información y análisis de datos.
- Empoderamiento del ciudadano a través de un Estado más abierto.
- Gestionar el ciclo de vida y la calidad de la información.
- Diseñar y publicar servicios de información orientados a nuestros grupos de interés.
- Analizar y visualizar la información a través de software libre.

▶ ¿Por qué abrir los datos?



El futuro digital
es de todos

MinTIC

LEY 1712 DE 2014. Ley de Transparencia y Acceso a la Información.
Conpes 3920 de 2018 – Política Nacional de Exportación de Datos BIG DATA
Política de Gobierno Digital – Guía Nacional de Datos Abiertos – Guía de Estándares de Calidad

Para acercar la **percepción** ciudadana a la realidad

Para evidenciar la realidad clara y oportunamente al ciudadano

Para dar a conocer logros de una administración

Para disminuir las PQR de los ciudadanos y mejorar la Calidad de mis datos

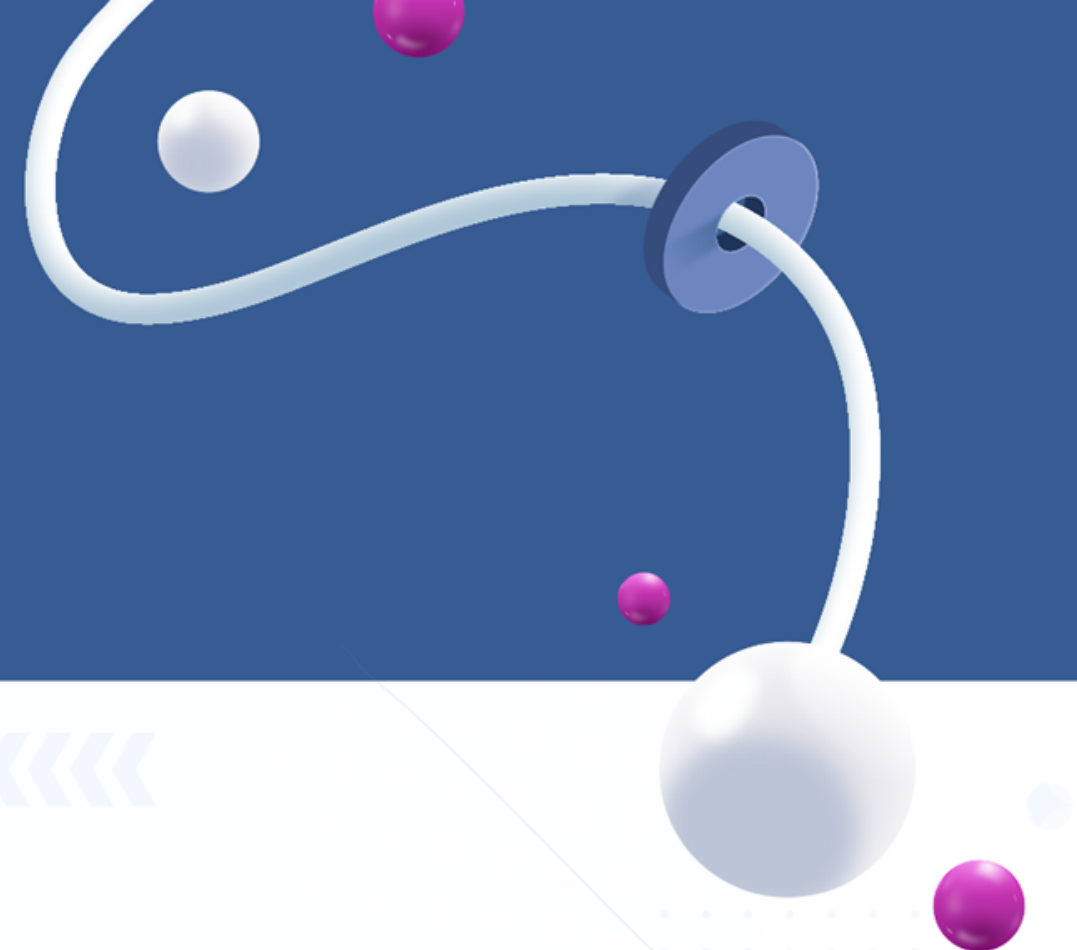
Apoyar el cumplimiento de los planes de desarrollo y objetivos estratégicos



El futuro digital
es de todos

MinTIC

MÓDULO 1



► Módulo 1: Calidad del Dato

¿Qué es calidad del dato? Generalidades

Calidad de datos se refiere al conjunto de procesos, técnicas, algoritmos y procedimiento encaminados a conseguir que los datos sean útiles para el fin que se le quiera dar en el negocio.



Que se busca?

- Determinar si los datos proporcionan una visión única de la temática a tratar.
- Que sean datos exactos, completos, consistentes, creíbles y actualizados.
- Que cumplan con las normativas o aspectos técnicos.
- La calidad de los datos depende de los consumidores, si estos pueden utilizarlos adecuadamente sin tener que depurarlos, limpiarlos o modificarlos se dice que los datos generan valor.



El futuro digital
es de todos

MinTIC

▶ Características y Criterios

Categorías



El futuro digital
es de todos

MinTIC

1. Calidad de Datos Inherentes

Grado con el que las características de calidad de los datos tienen el potencial intrínseco para satisfacer las necesidades establecidas y necesarias cuando los datos son utilizados bajo condiciones específicas.

Hace referencia a:

- Valores de dominios de datos y posibles restricciones (Reglas de Negocio con la calidad requerida por las características en una aplicación dada).
- Relaciones entre valores de datos (Consistencia).
- Metadatos.

2. Calidad del Dato Dependiente del Sistema

Grado con el que la Calidad de Datos es alcanzada y preservada a través de un sistema informático cuando los datos son utilizados bajo condiciones específicas.

Desde el punto de vista dependiente del sistema, la Calidad de Datos depende del dominio tecnológico en el que los datos se utilizan, y se alcanza mediante las capacidades de los componentes del sistema informático tales como: dispositivos hardware (e.g., Respaldo Software para alcanzar la Recuperabilidad), y otro software (e.g., Herramientas de migración para alcanzar la Portabilidad).

Exactitud, completitud, conformidad, consistencia

Criterios



El futuro digital
es de todos

MinTIC

Antes

Name	Gender	Street	House #	Zip code	City	State	D.O.B
John Doe	Male	60th Street	45		New York	New York	08/12/64
Jane Doe	Female	Jonathan	36	10023	Poughkeepsy	NY	21-dec-1954

Después

Name	Gender	Street	House #	Zip code	City	State	D.O.B
John Doe	Male	E 60th St	45	10022	New York	NY	08/12/64
Jane Doe	Female	Jonathan St	36	10023	Poughkeepsie	NY	21/12/54

- Exactitud
- Completitud
- Conformidad
- Consistencia

Exactitud: Datos diligenciados correctamente

Completitud: Información completa, datos completos para todas las columnas

Conformidad: cumplimiento de lineamientos y estándares vigentes para la descripción de la Metadata.

Consistencia: Datos coherentes y libres de contradicción

Crerios



Confidencialidad, relevancia, actualidad, trazabilidad, precisi3n



El futuro digital
es de todos

MinTIC

ID	Nombre_Cliente	Tipo_persona	Ultima_fecha_facturacion	Estatus_cliente	Direccion_1	Direccion_2	Direccion_3
541	Juan Luis Valdez	Persona	23-Jan-12	Activo	Av. Del Corral #45	Col. San Juan	Monterrey, N.L.
542	IBM	Empresa	11-Jul-12	Activo	Periferico Sur 39 Int.41	Col. Chapultepec	Mexico, D.F.
543	Andrea Sanchez	Persona		Activo	Av. Hidalgo No. 129	Col. Chapulin	Gdl.
544	Jose Luis Rosas	Persona	23-May-12	Baja	24F	Del. Juarez	Taxco
545	Coca-Cola Corp.	Empresa	6-Feb-13	Activo	Juan Polainas 253	Col. Alta Mata	Veracruz
546	Baby Crazy	Persona	23-Mar-12	Activo	Calle 4 No. 78	Del. Alvaro Obregon	Mexico, D.F.
547	Emilio Contreras Chavez	Empresa	18-Apr-12	Activo	Av. Insurgentes 31	Col. Escutia	Puebla
548	Felipe Calderon del Rio	Persona	11-Jun-13	Activo	C/O Juan de la Barrera	Col. Pedrito	San Luis Chalco, Edo. Mexico
549	Maria del Corral de Garduno	Persona	30-Sep-12	Activo	Av. 7 #428 Paseo de la Reforma No. 32	Col. San Sergio	Mexico, D.F.
550	OXXO	Empresa	4-Oct-12	Activo	Tollocan 52	Col. Cuahutemoc	Mexico, D.F.
551	Karina Salgado Marmolejo	Persona	3-Dec-12	Baja	Xochitecatl 295	Diamante	Acapulco, Guerrero
552	Julian Garduno Flores	Persona	1-Feb-13	Activo	Xochitecatl 295	Del. Xochimilco	Mexico, D.F.
553	Juan Valdes	Persona	23-Nov-12	Activo	Av. Del Corral #45	Col. San Juan	Monterrey, N.L.



Completitud: Falta de fecha para Andrea S3nchez

conformidad: Las direcciones no est3n de acuerdo a los est3ndares

Consistencia: Baby Crazy es empresa y Emilio Contreras en persona

Duplicaci3n

Confidencialidad: los datos solo deben ser accedidos por personas autorizadas para proteger adecuadamente la informaci3n reservada y clasificada.

Relevancia: los datos publicados deben ser de utilidad para los usuarios.

Actualidad: vigencia y actualidad de los datos publicados.

Trazabilidad: hist3rico del conjunto de datos disponible: fechas de creaci3n, publicaci3n y actualizaciones.

Precisi3n: Nivel de desagregaci3n en que est3n publicados los datos con respecto al nivel de desagregaci3n en que fueron generados

Criterios



Portabilidad, credibilidad, comprensibilidad, accesibilidad

Portabilidad: Formatos sin restricciones para la reutilización de los datos.

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.



El futuro digital
es de todos

MinTIC





▶ Criterios de calidad e interoperabilidad

- ✓ Confidencialidad
- ✓ Relevancia
- ✓ Actualidad
- ✓ Trazabilidad
- ✓ Conformidad
- ▶ Exactitud
- ▶ Compleitud
- ▶ Consistencia
- ▶ Precisión
- ✓ Portabilidad
- ✓ Credibilidad
- ✓ Comprensibilidad
- ✓ Accesibilidad



Pasar a ver ejemplos de cada uno

**MÁS
COMUNES**





Plan Calidad del Dato

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 para mejorar la calidad
- Realizar pruebas unitarias y funcionales a los datos ajustados

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 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.



Plan Calidad del Dato

Ejecución de la calidad



El futuro digital
es de todos

MinTIC



Semanal

El conjunto de datos contiene menos de cincuenta 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 sola columna ?
- Carga de archivos no admitidos como Datos Abiertos ?



Quincenal

Se están ajustando de forma automática datos como municipio, departamento, entidad y formato de fechas.

- El conjunto de datos no se encuentra actualizado a la fecha, de acuerdo al campo de la Metadata con la 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 ?
- Conjunto con columnas de solo texto ?
- Conjunto de datos no tiene filas con información ?



Herramientas

para la calidad del Dato



El futuro digital
es de todos

MinTIC

Herramientas

- Informática® Data Quality™
- Ataccama - DQ Analyzer, Data Quality Center
- IBM InfoSphere Information Server
- Neopost - Data Cleaner, Data Hub y Data Cloud
- Oracle Enterprise Data Quality
- Open Refine

Características

- Análisis de datos para capturar estadísticas de los metadatos
- Descomposición y transformación de campos de texto y contenido
- Modificación de los valores de datos para cumplir con restricciones de dominio
- Aprendizaje automático, base de conocimiento
- Interfaces usables disponible para usuarios no técnicos y técnicos



OpenRefine



LEILA

https://colaboracion.dnp.gov.co/CDT/Desarrollo%20Digital/Big%20Data/2020/00_LEILA/LEILA_Presentacion.pdf



▶ Qué no publicar / Qué publicar



El futuro digital
es de todos

MinTIC

Ejemplos para mejorar

NO

- Índice de información clasificada.
- Listado de restaurantes.
- IBR – Banco de la Nación
- Procesos Judiciales
- Zonas WIFI

SI

- Índice de información clasificada de la Alcaldía de Chía.
- Listado georreferenciado restaurantes municipio de Zipaquirá, Cundinamarca.
- Indicador Bancario de Referencia – Serie histórica
- Procesos Judiciales en contra del municipio de Soacha
- Zonas Wifi de Colombia



Qué no publicar / Qué publicar

Recomendaciones generales



El futuro digital
es de todos

MinTIC



- La escritura normal utiliza habitualmente las letras minúsculas.
- Las letras mayúsculas deben escribirse con tilde, si les corresponde.
- La información de título y descripción debe ser complementaria y dar en pocas palabras datos relevantes sobre la información a consultar.
- Las siglas se deben usar siempre y cuando se explique de qué se trata.



El futuro digital
es de todos

MinTIC



**Usuario publicador y
calidad de la metadata**

Publicación de datos

www.datos.gov.co

▶ Usuario publicador



El futuro digital es de todos

MinTIC

Inicio Descubre Publica Visualiza Participa Herramientas Novedades Usos Activos Calidad Español Iniciar sesión

Iniciar sesión a la plataforma de datos abiertos del gobierno colombiano

¿Ha olvidado su contraseña?

Iniciar sesión

¿Todavía no tiene una cuenta? [Registrarse](#)

GOV.CO

Políticas de Privacidad y Condiciones de uso [Soporte](#)

© 2020 Ministerio de Tecnologías de la Información y las Comunicaciones Edificio Murillo Toro Cra. 8a entre calles 12 y 13, Bogotá, Colombia - Código Postal 111711



www.datos.gov.co Catálogo de datos Crear Extra

Conjunto de datos

Historia

tdpt

Mis activos Compartidos conmigo

Tipo	Nombre	Acciones	Último actualizado	Categoría	Propietario	Des
	xxx ADVERTENCIA DE RESPONSABILIDAD: La variable etnia d...	...	September 19, 2020	Salud y Protección Social	tdpt	...
	Visualización datos Covid Municipio XX	...	September 18, 2020	Salud y Protección Social	tdpt	...



Crear nuevo conjunto de datos

Para empezar, elige un nombre para el conjunto de datos

Crear conjunto de datos

▶ Usuario publicador



El futuro digital
es de todos

MinTIC

Datos 1

Acciones de datos

+
Añadir datos
Añada datos tabulares, geoespaciales, no analizables o vinculados externamente
[Añadir datos](#)

☰
Revisar y configurar datos
[Revisar datos](#)

✎
Editar los metadatos del conjunto de datos
Añada un contexto para ayudar a los usuarios a comprender sus datos
[Editar metadatos](#)

✎
Editar columna de metadatos
[Editar metadatos](#)

Sobre este conjunto de datos



Arrastrar un nuevo archivo de datos aquí
...o pulse Explorar para elegir un archivo del ordenador

[Navegar](#)

Los tipos de archivos compatibles son: Valores separados por coma (CSV), Valores separados por tabulador (TSV), Microsoft Excel (XLS), Microsoft Excel (OpenXML), Archivo ZIP (shapefile), Formato JSON (GeoJSON), Formato GeoJSON, Lenguaje de marcado (KML), Lenguaje de marcado comprimido (KMZ)

Los archivos no analizables también pueden ser cargados para su almacenamiento

Las cargas de directorios no son compatibles



Vista previa de los datos (Guion_video tutorial_SetDatos_M1_UT2.csv)

ID	FechaMuestra	ZonaGeografica	TipoDocumento	Documento
#	Número	Fecha y hora	T Texto	# Número
1	2020 Mar 26 12:00:00 AM	BOGOTA DC-BOGOTA DC	Cedula de ciudadanía	1,234,567,891
2	2020 Mar 29 12:00:00 AM	BOGOTA DC-BOGOTA DC	Cedula de ciudadanía	1,234,567,892
3	2020 Apr 03 12:00:00 AM	BOGOTA DC-BOGOTA DC	Cedula de ciudadanía	1,234,567,893
4	2020 Mar 31 12:00:00 AM	BOGOTA DC-BOGOTA DC	Cedula de ciudadanía	1,234,567,894
5	2020 Mar 17 12:00:00 AM	BOGOTA DC-BOGOTA DC	Cedula de ciudadanía	1,234,567,895
6	2020 Apr 03 12:00:00 AM	BOGOTA DC-BOGOTA DC	Cedula de ciudadanía	1,234,567,896
7	2020 Mar 30 12:00:00 AM	BOGOTA DC-BOGOTA DC	Cedula de ciudadanía	1,234,567,897
8	2020 Mar 19 12:00:00 AM	BOGOTA DC-BOGOTA DC	Cedula de ciudadanía	1,234,567,898
9	2020 Mar 23 12:00:00 AM	BOGOTA DC-BOGOTA DC	Cedula de ciudadanía	1,234,567,899
10	2020 Apr 01 12:00:00 AM	BOGOTA DC-BOGOTA DC	Cedula de ciudadanía	1,234,567,900
11	2020 Apr 07 12:00:00 AM	BOGOTA DC-BOGOTA DC	Cedula de ciudadanía	1,234,567,901
12	2020 Mar 30 12:00:00 AM	BOGOTA DC-BOGOTA DC	Cedula de ciudadanía	1,234,567,902
13	2020 Apr 03 12:00:00 AM	BOGOTA DC-BOGOTA DC	Cedula de ciudadanía	1,234,567,903
14	2020 Mar 18 12:00:00 AM	BOGOTA DC-BOGOTA DC	Cedula de ciudadanía	1,234,567,904
15	2020 Mar 26 12:00:00 AM	BOGOTA DC-BOGOTA DC	Cedula de ciudadanía	1,234,567,905

▶ Calidad de los Metadatos



El futuro digital
es de todos

MinTIC

Acciones de datos

 Añadir datos Añada datos tabulares, geoespaciales, no analizables o vinculados externamente. Añadir datos	 Revisar y configurar datos Configure y confirme sus datos antes de publicar. Revisar datos	 Editar los metadatos del conjunto de datos Añada un contexto para ayudar a los usuarios a comprender sus datos. Editar metadatos	 Editar columna de metadatos Describe sus columnas para ayudar a los usuarios a interpretar sus datos. Editar metadatos
--	---	---	---

Sobre este conjunto de datos

Título y descripción
Haga su título y descripción tan claros y simples como pueda.

Título de Conjunto de Datos *

Breve Descripción

Etiqueta de la fila
Describe lo que represente cada fila (en caso aplicable).

Título de Fila

Categorías y etiquetas
Categorice su conjunto de datos para que sea más fácil encontrarlo.

Categoría

▶ Calidad de los Metadatos



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

<p>Actualizado 24 de agosto de 2021</p> <p>Última actualización de metadatos 24 de agosto de 2021</p> <p>Fecha de cración 24 de agosto de 2021</p> <hr/> <p>Vistas Descargas 4 0</p> <hr/> <p>Datos ofrecidos por Superintendencia de Sociedades Propietario de conjunto de datos amorales</p> <p>Contactar con dueño de conjunto de datos</p>	<p>Información de la Entidad</p> <table border="1"><tr><td>Área o dependencia</td><td>SUPERINTENDENCIA DE SOCIEDADES</td></tr><tr><td>Nombre de la Entidad</td><td>Superintendencia de Sociedades</td></tr><tr><td>Departamento</td><td>Bogotá D.C.</td></tr><tr><td>Municipio</td><td>Bogotá D.C.</td></tr><tr><td>Orden</td><td>Nacional</td></tr><tr><td>Sector</td><td>Comercio, Industria y Turismo</td></tr></table> <p>Información de Datos</p> <table border="1"><tr><td>Idioma</td><td>Español</td></tr><tr><td>Cobertura Geográfica</td><td>Nacional</td></tr><tr><td>Frecuencia de Actualización</td><td>Anual</td></tr><tr><td>Fecha Emisión (aaaa-mm-dd)</td><td>2021-08-24</td></tr><tr><td>URL Documentación</td><td>https://siis.ia.supersociedades.gov.co/#/</td></tr><tr><td>URL Normativa</td><td>http://www.supersociedades.gov.co/delegatura_aec/informes_empresariales/Paginas/sirfin.aspx</td></tr></table>	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	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
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																								
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																								



El futuro digital
es de todos

MinTIC



PREGUNTAS



El futuro digital
es de todos

MinTIC

¡MUCHAS GRACIAS!

Ing. ALIRIO HUMBERTO OTÁLORA VIVARES
Científico de Datos

