



GUÍA PARA APERTURA DE DATOS ABIERTOS

OFICINA DE TECNOLOGÍAS Y SISTEMAS DE INFORMACIÓN

Contenido

INTRODUCCION	2
1. OBJETIVO	2
2. ALCANCE	2
3. TÉRMINOS Y DEFINICIONES	3
4. MARCO NORMATIVO	4
5. DEFINICIÓN DE DATOS ABIERTOS	4
5.1. FORMATOS	5
5.2. LICENCIAMIENTO	5
5.3. PRINCIPIOS	6
5.4. CARACTERÍSTICAS DE LOS DATOS ABIERTOS	6
5.5. VENTAJAS DE LOS DATOS ABIERTOS	7
6. METODOLOGÍA APERTURA DATOS ABIERTOS	7
6.1. FASE1 - ESTABLECER EL PLAN DE APERTURA	7
6.2. FASE2 - ESTRUCTURAR Y PUBLICAR LOS DATOS	10
6.3. FASE3 - COMUNICAR Y PROMOVER EL USO DE LOS DATOS	13
6.4. FASE4 - MONITOREAR LA CALIDAD Y EL USO	13

INTRODUCCIÓN

En Colombia, se han realizado importantes esfuerzos para impulsar la apertura de datos públicos para el acceso a la información pública que es un derecho fundamental, orientado a garantizar la participación democrática y la transparencia de la gestión pública, por lo cual se constituye como un mecanismo de control ciudadano de la actividad estatal.

El Ministerio de Ciencia, Tecnología e Innovación, en cumplimiento de la Ley 1712 de 2014 de Transparencia y Acceso a la Información Pública, y sus normas reglamentarias, a través del presente lineamiento de Datos Abiertos, busca definir la estrategia de apertura y reúso de datos, con el propósito de generar valor en lo económico, social, político, cultural, ambiental, y en general, en los distintos ámbitos de la sociedad.

El objetivo del presente documento es proporcionar orientaciones, buenas prácticas y toma de decisiones basadas en datos a partir del aumento en el uso y aprovechamiento de la información para el desarrollo de la estrategia de apertura de datos de la entidad, para la puesta en práctica de datos abiertos brindados a la ciudadanía, a través de los medios establecidos.

Los datos abiertos son información pública dispuesta en formatos que permiten su uso y reutilización bajo licencia abierta y sin restricciones legales para su aprovechamiento. En Colombia, la Ley 1712 de 2014 sobre Transparencia y Acceso a la Información Pública Nacional, define los datos abiertos en el numeral sexto como “todos aquellos datos primarios o sin procesar, que se encuentran en formatos estándar e interoperables que facilitan su acceso y reutilización, los cuales están bajo la custodia de las entidades públicas o privadas que cumplen con funciones públicas y que son puestos a disposición de cualquier ciudadano, de forma libre y sin restricciones, con el fin de que terceros puedan reutilizarlos y crear servicios derivados de los mismos”¹.

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.

1. OBJETIVO

Definir los lineamientos para la identificación, actualización, publicación, uso y aprovechamiento de los datos abiertos, priorizando datos estratégicos y requisitos de calidad, a fin de tener la documentación necesaria que deba ser aplicable por los servidores públicos y contratistas para promover la apertura de los datos del Ministerio de Ciencia, Tecnología e Innovación, con el propósito de suministrar y publicar de manera transparente y eficiente la información que genera la entidad, en los diferentes medios definidos para conocimiento de las partes interesadas y ciudadanía en general y cumplir con lo establecido en la estrategia de Gobierno Digital (antes Gobierno en Línea).

2. ALCANCE

Presentar los lineamientos que se deben tener en cuenta sobre la iniciativa de Datos Abiertos que busca que el Ministerio de Ciencia, Tecnología e Innovación, publique la información pertinente y de calidad en formatos estructurados a disposición de los usuarios para que ellos y las entidades la utilicen de diferentes formas.

¹ [Guía para el uso y aprovechamiento de Datos Abiertos en Colombia](#)

3. TÉRMINOS Y DEFINICIONES

Anonimizar: Proceso técnico que hace que la información de identificación personal no sea visible, logrando que no sean identificables los individuos a los que pertenece la información. Anonimizar permite ocultar o eliminar los datos o características que identifique plena o parcialmente a las personas, organizaciones o características individuales de la fuente de información de manera irreversible, de tal forma que impidan identificar a una persona u organización.

Conjunto de datos: Una unidad mínima de información sujeta a carga, publicación, transformación y descarga en la plataforma.

Datos enlazados (linked data): Método de publicación de datos estructurados para que puedan ser interconectados y más útiles. Se basa en tecnologías Web estándar, tales como HTTP, RDF y los URI, pero en vez de utilizarlos para servir páginas web para los lectores humanos, las extiende para compartir información de una manera que puede ser leída automáticamente por ordenadores. Esto permite que sean conectados y consultados datos de diferentes fuentes.

Datos tabulares: Se recomienda el formato CSV, en los cuales las cadenas de caracteres que conforman los datos se acomodan en filas separadas por saltos de línea y columnas separadas por comas.

Datos estructurados: Se recomienda el uso de los formatos JSON o XML cuya especificación se encuentra disponible abiertamente.

Datos espaciales: Se recomienda el uso de los formatos GeoPackage, SHP, GeoJSON, WMS, WFS, KML, GeoTIFF

Estándar: Es un conjunto de características y requisitos que se toman como referencia o modelo y son de uso repetitivo y uniforme. Para que sea un estándar debe haber sido construido a través de consenso y refleja la experiencia y las mejores prácticas en un área en particular.

Formatos abiertos: Son formatos de archivo que se pueden crear y manipular por cualquier software, libre de restricciones legales.

Licencia abierta: Es un acuerdo de licencia copyleft diseñado para permitir a los usuarios compartir, modificar y usar libremente una base de datos, manteniendo la misma libertad para los demás.

Metadato: Los metadatos son "datos sobre datos" - es decir, los datos que describen los aspectos básicos de un conjunto de datos, por ejemplo, cuándo se creó el conjunto de datos, cuál es la agencia responsable de la base de datos, el formato de los datos, etc.

Open Data (Datos Abiertos): Filosofía y práctica que persigue que determinados datos de los Gobiernos estén disponibles de forma libre a todo el mundo, sin restricciones de copyright, patentes u otros mecanismos de control.

Reutilización de datos: Producto que se elabora a partir de los datos públicos, puede ser una visualización, una aplicación web, un servicio, un cuadro de mandos, una noticia o una información, un dibujo, una gráfica dinámica, entre otras cosas.

Reutilizadores de datos: Aquellas personas que con los datos abiertos como materia prima elaboran productos o servicios, pueden ser tales como emprendedores, empresas, ONGs, periodistas, hackers cívicos, o cualquier persona que tenga conocimientos del tratamiento y la manipulación de los datos.

Usabilidad: Significa facilidad de uso, atributo de calidad, que identifica el grado en que un producto puede ser usado por determinados usuarios para lograr sus propósitos con eficacia, eficiencia y satisfacción en un contexto de uso específico.

4. MARCO NORMATIVO

Marco Normativo	Descripción
Ley 1474 de 2011	“Por la cual se dictan normas orientadas a fortalecer los mecanismos de prevención, investigación y sanción de actos de corrupción y la efectividad del control de la gestión pública”.
Ley 1712 de 2014	“Por medio de la cual se crea la Ley de Transparencia y del Derecho de Acceso a la Información Pública”.
Decreto 2573 de 2014	Por el cual se establecen los lineamientos, instrumentos y plazos de la estrategia gobierno en línea para garantizar el máximo aprovechamiento de las Tecnologías de la Información y las Comunicaciones, con el fin de contribuir con la construcción de un Estado abierto, más eficiente, más transparente y participativo y que preste mejores servicios con la colaboración de toda la sociedad”.
Decreto 1008 de 2018	Por el cual se establecen los lineamientos generales de la política de Gobierno Digital y se subroga el capítulo 1 del título 9 de la parte 2 del libro 2 del Decreto 1078 de 2015, Decreto Único Reglamentario del sector de Tecnologías de la Información y las Comunicaciones.
Ley 1581 de 2012	Protección de datos personales
Decreto Reglamentario Único 1081 de 2015 Decreto 103 de 2015	Este decreto tiene por objeto reglamentar la Ley 1712 de 2014, en lo relativo a la gestión de la información pública.
Tutorial para la gestión de usuarios y creación	Tutorial de la plataforma www.datos.gov.co de MinTIC
CONPES 3920 DE 2018	Política Nacional de Explotación de Datos (Big Data)
CONPES 3975 DE 2018	Política Nacional para la Transformación Digital e Inteligencia Artificial
Guía para el uso y aprovechamiento de Datos Abiertos en Colombia	Orientaciones y buenas prácticas para el desarrollo de estrategias de apertura y uso de datos www.gobiernodigital.gov.co MinTIC
Resolución 1519 de 2020	Por la cual se definen los estándares y directrices para publicar la información señalada en la Ley 1712 del 2014 y se definen los requisitos materia de acceso a la información pública, accesibilidad web, seguridad digital, y datos abiertos

5. DEFINICIÓN DE DATOS ABIERTOS.

En Colombia la Ley de Transparencia y Acceso a la Información, define los datos abiertos como “todos aquellos datos primarios o sin procesar, que se encuentran en formatos estándar e interoperables que facilitan su acceso y reutilización, los cuales están bajo la custodia de las entidades públicas o privadas que cumplen con funciones públicas y que son puestos a disposición de cualquier ciudadano, de forma libre y sin restricciones, con el fin de que terceros puedan reutilizarlos y crear servicios derivados de los mismos” (Ley 1712 de 2014. Literal J, artículo 6. Definiciones.)

De este modo, la Ley establece la obligatoriedad de las entidades públicas de “divulgar datos abiertos”, teniendo en cuenta las excepciones de acceso a la información, asociadas a información clasificada y reservada establecidas en su título tercero, artículos 18 y 19. A continuación, se señalan algunos ejemplos de datos abiertos de entidades públicas disponibles en el Portal de Datos del Estado Colombiano:

Las contrataciones públicas dispuestas por la Agencia de Contratación - Colombia Compra Eficiente, donde se disponen todos los gastos en contratación que realiza el estado, permitiendo saber a quién le compra el Estado, cuanto gastó y en qué gasto.

Se debe tener en cuenta que la apertura de datos va más allá de la preparación y publicación de los datos por parte de las entidades, también involucra acciones concretas como promover el uso de los datos publicados, realizar seguimiento e identificar el valor agregado que se está generando con su utilización.

Además, en el proceso de apertura y uso de los datos, se debe monitorear la calidad y el uso, la estandarización de la calidad y mejorar la interoperabilidad de los datos abiertos publicados o por publicar, en los portales dispuestos por las entidades nacionales o distritales.

Los datos abiertos deben ser publicados en sitios web de manera gratuita, accesibles al público, dispuestos en formatos que permiten su uso, reutilización y aprovechamiento sin restricciones legales y bajo licencia abierta.

5.1. FORMATOS.

Algunos de los formatos de datos abiertos más comunes son:

- **CSV (Valores Separados por Coma):** es un tipo de documento de texto plano para representar datos tabulados en columnas separadas por comas (o punto y coma) y filas separadas por saltos de línea.
- **XML (Lenguaje Etiquetado Extensible):** es el estándar utilizado para el intercambio de información estructurada entre diferentes plataformas.
- **RDF (Infraestructura para Descripción de Recursos):** es un lenguaje de propósito general para representar información en la web, es una de las tecnologías esenciales para la web semántica.
- **RSS (Sindicación Realmente Sencilla):** es un formato basado en XML para la ordenación de contenidos de páginas web actualizados frecuentemente (noticias, blogs o podcasts).
- **JSON (Notación de objetos Javascript):** es un formato simple y ligero para el intercambio de datos, de fácil lectura e interpretación por las personas y sencillo para ser analizado y generado por las máquinas.
- **ODF (Formato de Documento Abierto para Aplicaciones Ofimáticas):** es un formato de fichero estándar para el almacenamiento de documentos ofimáticos.
- **WMS (Servicio Web de Mapas):** es un servicio definido por el OGC (Open Geospatial Consortium) que produce mapas de datos referenciados espacialmente, de forma dinámica a partir de información geográfica.
- **WFS (Servicio Web de características):** es un servicio estándar, que ofrece una interfaz de comunicación que permite interactuar con los mapas servidos por el estándar WMS.

5.2. LICENCIAMIENTO

Las licencias estándar en el entorno internacional se han constituido en un elemento de confianza digital en el uso de datos abiertos, ya que están concebidas como una forma de protección de los derechos de autor y de apoyo al dominio público de las obras protegidas, fomentando el acceso a la información pública, orientaciones sobre licencias pueden ser identificadas en el instructivo para la definición de licencias de datos, elaborado por la Unidad Administrativa Especial de Catastro Distrital - IDECA².

² Unidad Administrativa Especial de Catastro Distrital IDECA. Instructivo para la definición de licencias de datos. Recuperado el 22 de 4 de 2019, en <https://www.ideca.gov.co/sites/default/files/documentacion/instructivolicencadatos.pdf>.

La licencia abierta Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0)³ permite hacer uso, transformación, distribución, redistribución, reutilización, compilación, extracción, copia, difusión, modificación y/o adaptación de los datos abiertos, citando siempre la fuente, este tipo de licencia reconoce la producción o generación de los datos por la entidad publicadora, señalando que ésta no será responsable de la utilización o transformación de los mismos datos, ni tampoco de daños sufridos o pérdidas económicas que, de forma directa o indirecta, produzcan o puedan producir perjuicios económicos, materiales o sobre datos y aplicaciones, provocados por el uso y/o transformación de los datos, esta manifestación regularmente la adopta la entidad en la plataforma de datos donde los dispone, en Colombia se encuentra en el Portal de Datos del Estado Colombiano.

5.3. PRINCIPIOS

Los principios⁴ y buenas prácticas para la liberación de datos abiertos han sido establecidos en la Carta Internacional de Datos Abiertos, adoptada por Colombia desde 2016 como instrumento orientador en la generación y uso de datos. Para que un dato sea considerado útil debe contar con ciertas características:

- ✓ **Abiertos por defecto:** La información debe estar disponible para la consulta ciudadana desde el momento de su generación.
- ✓ **Comparables e interoperables:** Pueden ser comparables desde distintos sectores territorios y periodos de tiempo.
- ✓ **Oportunos y completos:** Los datos abiertos son relevantes solo si aportan valor para el usuario de la información.
- ✓ **Mejorar la gobernanza y participación ciudadana:** Abrir datos fortalece la gobernanza y favorece la transparencia.
- ✓ **Accesible y utilizables:** Son gratuitos y bajo licencia abierta, facilitando los datos en la mayor cantidad de usuarios y maquinas.
- ✓ **Apoyar el desarrollo incluyente y la innovación:** El uso de datos abiertos permite la construcción de nuevo conocimiento que puede ser usado en la generación de valor social y económico.

5.4. CARACTERÍSTICAS DE LOS DATOS ABIERTOS

Los datos del Gobierno son considerados abiertos si son publicados siguiendo los siguientes principios o características:

- ✓ **Primarios:** Los datos deben obtenerse en la fuente de origen, con el más alto nivel de detalle posible, no en forma agregada ni modificada.
- ✓ **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 encontrarse en formatos que permitan el procesamiento automático.
- ✓ **No propietarios:** Los datos deben estar disponibles en un formato sobre el cual ninguna entidad tenga control exclusivo.
- ✓ **No discriminados:** Los datos deben estar disponibles para cualquiera persona, sin requerir registro o autenticación.
- ✓ **Licenciados de forma abierta:** Los conjuntos de datos publicados deben contar con términos de uso y licenciamiento abierto, es decir, no deben estar sujetos a ninguna normativa de derecho de autor, patente, marca registrada o regulaciones de acuerdo de secreto. Se podrán permitir restricciones razonables de privacidad, seguridad o privilegios según lo regulado por otros estatutos.
- ✓ **Completos:** Los datos públicos deben reflejar la totalidad del tema y contener el mayor detalle posible, garantizando que la información suministrada sea suficiente y consistente y que no contenga datos nulos. Además, no deben estar

³ Commons, C. (s.f.). Attribution-ShareAlike 4.0 International (CC BY-SA 4.0). Recuperado el 22 de 4 de 2019, en <https://creativecommons.org/licenses/by-sa/4.0/deed.es>

⁴ Guía para el uso y aprovechamiento de Datos Abiertos en Colombia pag.7

sujetos a privacidad u otras limitaciones y estar electrónicamente almacenados incluyendo, pero no limitado a documentos, bases de datos, transcripciones y grabaciones audiovisuales.

- ✓ **Oportunos:** Debe estar disponible tan rápido como sea necesario para garantizar su valor y mantener una frecuencia de actualización que garantice la utilidad del dato.

5.5. VENTAJAS DE LOS DATOS ABIERTOS

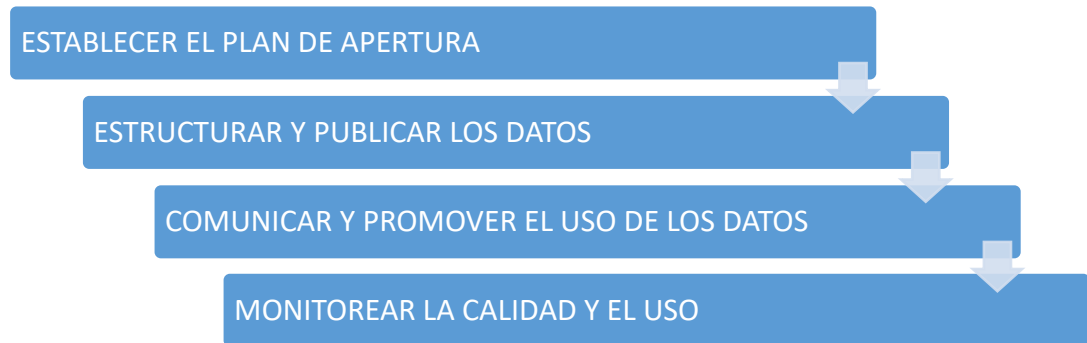
Al reutilizar con efectividad los Datos Abiertos, estos pueden generar beneficios desde las perspectivas económicas y sociales, desarrollando procesos que permitan generar valor en áreas como la transparencia y control social, mejorando o creando productos o servicios.

A continuación, se presentan algunos ejemplos de estos beneficios:

- ✓ Disponer de la información pública del quehacer de las entidades en datos abiertos, permite que la sociedad civil, las entidades públicas y otros grupos de interés, realicen procesos de investigación y control social a sus gobernantes.
- ✓ Las entidades públicas mejoran la prestación de servicios a los ciudadanos, así como a optimizar procesos internos.
- ✓ Al analizar los datos estos permiten pronosticar y prevenir fenómenos, mejorando la toma de decisiones.
- ✓ Son insumo fundamental para la generación de nuevo conocimiento.

6. METODOLOGÍA APERTURA DATOS ABIERTOS

Para conseguir la apertura de los datos del Ministerio, se propone la siguiente metodología compuesta por las fases que se presentan en la siguiente imagen:



6.1. FASE1 - ESTABLECER EL PLAN DE APERTURA

De acuerdo con el marco normativo vigente, como primer paso es necesario adelantar una estrategia para el proceso de apertura, identificar los actores que harán parte del ecosistema tanto publicadores con reutilizado res, usar las herramientas disponibles para apertura, calidad y uso de sus datos abiertos, diseñar un esquema de incentivos y generar estrategias de monitoreo al valor de los datos abiertos.

En la planificación se generan las directrices y se define el marco de gobernanza de la iniciativa de datos a planificar, esto incluye:

- ✓ Identificación de las tareas a realizar
- ✓ Recursos
- ✓ Tiempos
- ✓ Productos que se obtendrán
- ✓ Esfuerzo que se destinará para las diferentes tareas
- ✓ Impacto esperado de los datos a abrir

Para ello es importante trazarse unos objetivos claros que respondan a las preguntas claves en un proceso de planeación respondiendo a ¿Qué?, ¿Para qué?, ¿Cómo?, ¿Por qué?, ¿Con qué?, ¿Cuándo?, ¿Con quién?

El plan de apertura de datos abiertos incluye las siguientes actividades:



a. Identificar: La identificación busca realizar un levantamiento de la información pública que administra la entidad y que puede llegar a convertirse en datos abiertos. Para efectos de realizar la identificación de esta información y de conformidad con lo establecido en la normatividad vigente, se debe tener en cuenta durante esta fase los siguientes datos:

- ✓ Los datos de los registros de activos de información oficiales generados por la entidad clasificados como dato abierto.
- ✓ Los datos públicos generados o administrados por los Sistemas de Información de la entidad.
- ✓ Los datos públicos utilizados por herramientas digitales, aplicativos web o móviles desarrollados de la entidad.
- ✓ Los datos solicitados por la ciudadanía de manera recurrente mediante solicitudes de acceso a la información o los que deriven de ellas; también aquellos que puedan resolver alguna problemática de la entidad.
- ✓ Los datos de otras entidades o de dependencias al interior de la misma entidad con el fin de evitar la duplicidad a la hora de publicar datos en el portal. Para ver el inventario de todos los datos abiertos existentes en el Portal de Datos del Estado Colombiano consulte: <https://www.datos.gov.co/dataset/Vista-Inventario-de-Datos-Portal-Nacional-de-Datos/5vw4-mas9>

Para este momento, es importante vincular a los actores del ecosistema de datos, para que desde la apertura de los datos, se atiendan las necesidades específicas de la ciudadanía y partes interesadas, para darle continuidad al ciclo de datos y garantizar la pertinencia y oportunidad de los mismos.

De esta manera, se contribuye a promover la reutilización de datos de la entidad y difundir soluciones desarrolladas con datos abiertos, para ello se pueden vincular los actores del ecosistema mediante la ejecución de las siguientes actividades:

- ✓ Desarrollar procesos de co-creación de herramientas tecnológicas o aplicaciones para identificar conjuntos de datos que generarían valor al disponerlos en formato abierto, con participación de desarrolladores y usuarios.
- ✓ Realizar talleres o encuentros con usuarios o grupos de interés para generar propuestas de creación de conjuntos y usos creativos con los datos abiertos.
- ✓ Desarrollar planes o estrategias de incentivos para promover la participación de los empleados, usuarios y grupos de interés en el desarrollo de las actividades que impulsen la identificación y uso de datos abiertos de la entidad.

Una vez identificados los datos abiertos del Ministerio, se deben identificar las fuentes de información que generan dichos conjuntos de datos y determinar si la generación de estos se realiza de manera **manual** (*disponible en hojas de cálculo, por ejemplo*) o **automática** (*proviene de un sistema de información*).

Si la fuente de los datos es un sistema de información, se debe identificar y priorizar la adecuación y desarrollo de las funcionalidades requeridas en el sistema de información (software), para que de esta manera el proceso de carga de los datos abiertos sea automatizado a través de proceso de extracción, transformación y carga.

- b. Analizar:** Se analiza la información identificada en la primera fase del proceso, para determinar la información publicable y validar que no tenga ninguna restricción normativa bajo la Ley Colombiana.

Los datos abiertos no son contrarios a la protección de datos personales: si existe información asociada a datos personales que puede ser valiosa como dato abierto, se debe adelantar un proceso de **anonimización** para eliminar aquella información sensible que afecta la privacidad de personas u organizaciones.

Anonimizar los datos es un aspecto fundamental y obligatorio antes de la publicación de datos personales sensibles, este proceso técnico que hace que la información de identificación personal no sea visible, logrando que no sean identificables los individuos a los que pertenece la información.

Este proceso permite ocultar o eliminar los datos o características que identifique plena o parcialmente a las personas, organizaciones o características individuales de la fuente de información de manera irreversible, de tal forma que impidan identificar a una persona u organización.

En materia de datos abiertos, específicamente la Ley 1712 de 2014 de Transparencia y del derecho de acceso a la información pública establece que las entidades deben “publicar datos abiertos”, teniendo en cuenta las excepciones de **publicar información pública clasificada o información pública reservada**, por lo que en esta fase se hace necesario analizar los datos identificados en el marco de la legislación vigente.

- c. Priorizar:** La priorización de datos de valor es importante para que los esfuerzos de gobernanza, apertura y publicación estén enfocados en la provisión de datos de utilidad, valor y con potencial de crear impacto en la ciudadanía y de solucionar problemáticas públicas. Por lo anterior, se recomienda tener en cuenta los siguientes criterios a la hora de realizar la priorización de datos abiertos que serán incluidos en el Plan de Apertura:
- ✓ Los datos que permiten dar cumplimiento a metas institucionales de transparencia, participación, control social y colaboración.
 - ✓ Los datos que contribuyen a un propósito de política pública específico y que hagan parte de las categorías establecidas por las mejores prácticas a nivel nacional e internacional en la materia (Ver Anexo I: índices internacionales y marco normativo que fijan prioridades de publicación), y aquellos solicitados directamente por otras entidades públicas.
 - ✓ Los datos que favorecen al cumplimiento de las metas trazadas en el Plan Nacional de Desarrollo y los planes departamentales y municipales.
 - ✓ Los datos que aportan al cumplimiento de los Objetivos de Desarrollo Sostenible.

- ✓ Los datos abiertos de gran tamaño, alta concurrencia de usuarios y de permanente actualización, esto con el fin de que se planee la construcción de procesos automáticos de extracción, transformación y carga de datos (ETL) desde los sistemas de información de la entidad al Portal de Datos del Estado Colombiano www.datos.gov.co.
- ✓ Los datos que han sido solicitados por los ciudadanos en ejercicios virtuales o presenciales de consulta pública.

6.2. FASE2 - ESTRUCTURAR Y PUBLICAR LOS DATOS

Una vez se tiene el inventario de datos abiertos priorizados en el Plan de Apertura generado de la fase anterior, se requiere realizar una preparación para publicar los datos abiertos, la cual consiste en alistar los datos para que éstos puedan ser procesados y reutilizados por terceros.

Este alistamiento incluye tres actividades:



- a. **Documentar:** Los metadatos son características que describen los aspectos básicos de un conjunto de datos y son una herramienta fundamental para organizar, clasificar, relacionar y encontrar datos. La cantidad y calidad de metadatos asociados a un conjunto de datos, permite una mejor clasificación y búsqueda y recuperación de los datos.

Algunos ejemplos de metadatos son los siguientes:

- ✓ **Nombre del conjunto de datos:** título que describe de manera corta el conjunto de datos. Este debe ser suficientemente descriptivo para el ciudadano por lo cual se recomienda usar lenguaje claro y eliminar las siglas, caracteres especiales o abreviaturas del nombre del conjunto.
- ✓ **Descripción:** resumen descriptivo del conjunto de datos, en donde se explica el contenido, procedencia, contexto en tiempo y lugar, así como el detalle de lo que se encontrará, de manera que cualquier usuario se encuentre informado sobre el contenido antes de descargar el recurso.
- ✓ **Entidad:** nombre de la entidad o institución que produce o controla la información.
- ✓ **Categoría:** clasificación conceptual básica del conjunto de datos en sistemas de categorías o taxonomías disponibles tales como Agricultura, Transporte, Salud, Educación, etc.
- ✓ **Licencia:** licencia bajo la cual se publica el conjunto de datos. Ver tipos de licencias abiertas en <https://co.creativecommons.org/> y en <https://www.ideca.gov.co/sites/default/files/documentacion/instructivolicencadatos.pdf>.
- ✓ **Fecha de Publicación:** fecha en la que se publicó el conjunto de datos.
- ✓ **Periodicidad:** indica la frecuencia de actualización del conjunto de datos.

La cantidad y características de los metadatos, dependerá de la plataforma en donde serán publicados los conjuntos de datos. El Portal de Datos del Estado Colombiano <http://www.datos.gov.co/> usa el estándar de metadatos 7DCAT para describir los conjuntos de datos dentro de los catálogos publicados, lo que aumenta el potencial de ser encontrados en búsquedas en la web, y ayuda a que aplicaciones externas puedan consumir fácilmente los conjuntos de datos disponibles.

Para el caso de los datos geográficos frecuentemente se encuentran documentados a través de metadatos geográficos, que siguen referentes internacionales como la ISO19115 – “Geographic information – Metadata”. Este estándar contiene los elementos necesarios para facilitar el descubrimiento de datos y su acceso. Contiene además los elementos mínimos del metadato propuesto por DCAT y www.datos.gov.co.5

Adicionalmente, es necesario documentar cada dato a publicar en el conjunto de datos, esto se realiza mediante el estándar de Lenguaje Común de Intercambio de Datos, que permite representar el concepto teniendo en cuenta sus relaciones, validaciones y estructura.

- b. Estructurar:** Consiste en representar los datos para que se dispongan en formatos procesables por máquina y abiertos, asegurando que los datos estén disponibles de modo que puedan ser utilizados por computadoras y usuarios. La estructuración de datos busca transformar el dato de formatos no-textuales a formatos que permitan su manipulación y/o su conexión con otros datos (datos enlazados).

Existen diversos formatos para la estructuración de datos, pero a la hora de seleccionarlos se debe tener en cuenta que sean legibles directamente por máquinas, que sean abiertos, permitan un uso sin restricciones y se encuentren dentro de los más utilizados. No deben utilizarse formatos cerrados o no editables como por ejemplo información en formato PDF.

Los datos se desactualizan con el tiempo, por ello, las herramientas y procesos de calidad de datos actúan sobre los datos para corregir su deterioro. Existen diversos tipos de formatos para publicar datos abiertos, por lo cual, se recomienda seguir los siguientes formatos de publicación:

- ✓ **Datos tabulares:** Se recomienda el formato CSV, en los cuales las cadenas de caracteres que conforman los datos se acomodan en filas separadas por saltos de línea y columnas separadas por comas.
- ✓ **Datos estructurados:** Se recomienda el uso de los formatos JSON o XML cuya especificación se encuentra disponible abiertamente.
- ✓ **Datos espaciales:** Se recomienda el uso de los formatos GeoPackage, SHP, GeoJSON, WMS, WFS, KML, GeoTIFF

A fin de que los datos sean más útiles, estos deben ser fáciles de encontrar y comparar por diferentes variables como sectores, localización geográfica, fechas, entre otros, por lo cual, se deben emprender acciones para mejorarlos tales como:

1. Fomentar la interoperabilidad de los datos publicados mediante su estandarización, para conseguir esto las áreas generadoras o administradoras de Datos deberá aplicar:
 - ✓ **El Marco de Interoperabilidad** que le permitirá cumplir con los estándares establecidos para el intercambio de datos abiertos entre distintas entidades públicas.
 - ✓ **El Lenguaje Común de Intercambio de información** que es el estándar definido para intercambiar información entre las entidades, facilitando el entendimiento de los involucrados en los procesos de intercambio de información.
 - ✓ **Lineamientos técnicos para la producción y Gestión de información geográfica** para los datos abiertos geográficos, los cuales frecuentemente se encuentran en formatos que solo pueden ser legibles por software de Sistemas de Información Geográfica. Sin embargo, sin importar el tipo de licencia del software en el que se estructuraron los datos geográficos, estos deben ser dispuestos en formatos interoperables que permitan que cualquier usuario pueda hacer uso de ellos en un software libre.
2. Mejorar la usabilidad de los datos, en aspectos como la calidad, legibilidad y lógica de publicación de los conjuntos y recursos de datos. Para conseguir esto, es necesario tener en cuenta que:

⁵ Guía para el uso y aprovechamiento de Datos Abiertos en Colombia Pág. 22

- ✓ Los datos abiertos de alta calidad son una condición previa para analizarlos, reutilizarlos y garantizar su valor. 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.
- ✓ Es importante evaluar la calidad de los datos para verificar su exactitud, confiabilidad y aptitud para el uso. Para tal fin se puede hacer uso del guía de estándares de Calidad e Interoperabilidad de los Datos Abiertos del Gobierno de Colombia.
- ✓ En el caso de datos geográficos, se debe monitorear o evaluar su calidad antes de ser publicados. Para tal fin se puede hacer uso del estándar internacional ISO 19157 – “Geographic information - Data Quality” o de los instrumentos en idioma español que en Colombia existen para su adopción en www.icde.org.co.

Las evaluaciones de calidad más recomendadas son:

- ✓ Exactitud posicional (absoluta o relativa): evalúa que los datos de ubicación o posición coincidan con la realidad, teniendo en cuenta un sistema de referencia por coordenadas.
 - ✓ Totalidad: disponer de la totalidad de información espacialmente generada, atendiendo a eventuales restricciones en alguno(s) de sus atributos por las excepciones que dispone la Ley 1712 de 2014.
 - ✓ Consistencia de formato/Usabilidad: se mide el nivel de uso del paquete de datos geográficos a publicar, la funcionalidad y capacidad de accesibilidad a través de los portales que empleen de visitantes, es decir, que se encuentren en formatos interoperables.
- c. **Publicar:** Este paso o actividad consiste en poner a disposición de los usuarios el conjunto de datos abiertos a través del cargue de datos en una plataforma que permita la organización y fácil consulta por parte de quienes van a reutilizar los datos, de manera que se convierta en un sitio de acceso común y que tenga la posibilidad de hacer escalable la información publicada.

El Portal de Datos del Estado Colombiano www.datos.gov.co, es el sitio creado por el Ministerio TIC, en donde las entidades públicas del orden nacional y territorial deben publicar los datos abiertos de interés nacional en temas de su competencia. Este sitio cuenta con los formatos y procedimientos específicos para realizar la documentación estructuración, cargue y publicación de la información, busca promover la transparencia, el acceso a la información pública, la competitividad, el desarrollo económico, y la generación de impacto social a través de la apertura, la reutilización de los datos públicos, y el uso y apropiación de las TIC de acuerdo con la estrategia de Gobierno en Línea de Colombia.

Algunas herramientas para facilitar la publicación de datos abiertos son las siguientes:

- ✓ **Automatización de carga de datos:** este proceso se realiza a través de procesos de extracción, transformación y carga denominados ETL, por sus siglas en inglés, que permitirán ejecutar automáticamente el proceso de carga de los datos abiertos desde los sistemas de información de las entidades públicas a www.datos.gov.co con la periodicidad deseada y previamente configurada. Estos procesos se recomiendan cuando se deben cargar grandes volúmenes de datos con una alta periodicidad (tiempo real, diaria, semanal) y que se requiera dejar sistematizado el proceso de actualización de datos.
- ✓ **Federación de portales de datos abiertos:** este proceso consiste en referenciar o enlazar conjuntos de datos de portales externos de datos abiertos, de tal manera que se puedan acceder desde www.datos.gov.co, unificando el acceso a los datos e integrando las iniciativas locales de datos abiertos del sector público. Para llevar a cabo este

proceso es necesario que el portal de datos externo cuente con el estándar de metadatos DCAT y se realice el proceso que se describe en el manual de federación.

- ✓ **Carga manual de datos:** Este proceso se realiza a través de la herramienta de carga dispuesta en www.datos.gov.co, el ofrece tres opciones de carga.
 - Carga tradicional desde un archivo en los formatos .csv, tsv, .xls, .xlsx, .zip (shapefile), json (GeoJSON), geojson, kml, kmz, buscando el archivo fuente en su computadora.
 - Importar conjunto de datos tabular o geoespacial desde una URL que está alojada externamente para tipos de archivo en formatos .csv, tsv, .xls, .xlsx, .zip (shapefile), json (GeoJSON), geojson, kml, kmz)
 - Enlace a una fuente de datos externa: permite vincular a una fuente de datos externa para referenciar varios archivos a un conjunto de datos y vincular múltiples conjuntos de datos al origen de datos externo general.

Si su entidad reporta información a sistemas nacionales (Por ejemplo, SECOP, SIIF, SIMAT) no es necesario que publique esta información a título propio, dado que estas entidades están obligadas a publicar toda la información completa y consolidada del país.

a. FASE3 - COMUNICAR Y PROMOVER EL USO DE LOS DATOS

Los productores deben informar a sus audiencias sobre la publicación de los datos y establecer mecanismos que permitan identificar y difundir los productos, servicios, artículos periodísticos, investigaciones académicas y en general cualquier producto que use datos.

Es por eso por lo que los datos se pueden publicar en la sección de transparencia del Portal Institucional además del Portal Datos del Estado Colombiano www.datos.gov.co que busca promover la transparencia, el acceso a la información pública.

Adicionalmente, se recomienda promocionar la publicación de los conjuntos a través de las redes sociales del Ministerio.

b. FASE4 - MONITOREAR LA CALIDAD Y EL USO

El uso de los datos abiertos es un aspecto estrechamente ligado a su calidad. En la medida en que se garantice que los datos abiertos publicados son completos, oportunos y consistentes, es posible promover su uso y, por lo tanto, generar valor a partir de ello.

El equipo de trabajo a cargo en el Ministerio de Ciencia, Tecnología e Innovación debe adelantar acciones y establecer controles, que les permitan garantizar la calidad de los datos desde la fuente o recolección de los datos, para mantener altos estándares de calidad en el ciclo de vida de estos.

Para monitorear la calidad y el uso de los datos abiertos, se deben tener en cuenta las siguientes acciones de seguimiento:



a. Medir calidad

Para monitorear la calidad de los datos abiertos publicados, se ha construido un formato que comprende todos los criterios propuestos por MINTIC para realizar la medición a partir de indicadores de tipo cuantitativo que permiten establecer métricas para determinar su calidad.

Adicional a los propuestos en el formato, se propone complementar la medición con los siguientes:

- ✓ **Número de solicitudes de apertura de datos:** identificar las PQRDS relacionadas con acceso a la información y analizar en la sección participa del Portal de Datos <https://datos.gov.co/nominate> las solicitudes ciudadanas sobre sus temas, disponible para descargue de solicitudes, en el siguiente enlace: <https://www.datos.gov.co/Participaci-n-ciudadana/Datos-Abiertos-Solicitados-Par-Los-Ciudadanos-En-L/9it7-3gvu>
- ✓ **Número de dependencias o áreas estratégicas de la entidad que abren datos:** Es importante vincular a las dependencias de la entidad que contengan información estratégica y de alto valor e impacto al momento de la publicación.
- ✓ **Porcentaje de conjuntos de datos abiertos estratégicos publicados:** Luego del plan de apertura se debe hacer seguimiento a las metas de apertura establecidas de datos estratégicos de la entidad.
- ✓ **Número de conjuntos de datos descargados vs. número de conjuntos de datos publicados:** Para evaluar este indicador recomendamos realizar una vista filtrada del inventario de conjuntos de datos en www.datos.gov.co que le permita visualizar solo los conjuntos de datos de la entidad y validar las descargas por parte de los ciudadanos.
- ✓ **Número de conjuntos de datos visitados vs. descargados:** Este indicador permitirá identificar aquellos conjuntos que son visitados con recurrencia pero que el ciudadano decide no descargar, para revisar si el conjunto tiene temas de calidad por mejorar o comentarios ciudadanos solicitando mejoras puntuales.

b. Medir usos

Con el fin de mapear el impacto que está teniendo el uso y aprovechamiento de datos abiertos, se puso en marcha una sección de usos de datos abiertos, dispuesta en www.datos.gov.co que tiene como objetivo involucrar a quienes desarrollan productos (apps, páginas web, modelos predictivos, visualizaciones, periodismo de datos) para que divulguen su desarrollo.

Este espacio tiene como objetivo:

- ✓ Visibilizar y dar mayor alcance a los productos a partir de datos.
- ✓ Recopilar buenas prácticas en uso y aprovechamiento de datos.
- ✓ Crear comunidad entre quienes usan y aprovechan los datos.
- ✓ Identificar cuáles son los sectores /temáticas y tipos de productos que más se realizan.

Para monitorear los usos de los datos abiertos, es posible utilizar algunos indicadores de tipo cuantitativo que permiten determinar los productos desarrollados a partir de sus datos abiertos tales como:

- ✓ **Número de soluciones generadas a partir de datos abiertos:** Este indicador es uno de los más importantes porque permite identificar la usabilidad de los datos publicados y si se cumple con el impacto esperado al publicarlos, por lo cual, si el conjunto de datos no tiene asociados usos, ni descargas es un indicador importante que el dato no es útil y debe mejorarse o cambiarse.
- ✓ **Número de proyectos que hicieron uso de los conjuntos de datos abiertos:** entre estos pueden ser artículos de investigación y periodísticos, noticias, libros, visualizaciones, etc.
- ✓ **Número de proyectos de emprendimiento o nuevos negocios** generados a partir del uso de datos abiertos.

c. Medir impacto

Las actividades de monitoreo también deben orientarse a identificar el valor generado para ciudadanos, usuarios y grupos de interés a partir de la publicación de datos abiertos, con lo cual se apunta al fortalecimiento de las acciones en materia de Estado Abierto. En este sentido se recomienda:

- ✓ Realizar mediciones periódicas para todos los indicadores con el propósito de identificar tendencias sobre los datos de mayor impacto y tomar decisiones sobre acciones a seguir en el plan de apertura.
- ✓ Apoyar a las áreas estratégicas o entidades aliadas en el desarrollo de sus capacidades para recolectar, gestionar, analizar, medir y definir los datos más estratégicos que se van a aperturar, estableciendo responsabilidades claras en la gobernanza de la información y evitando duplicidad de datos publicados en una misma entidad o sector.
- ✓ Adaptar la estrategia de comunicación dirigida a los actores del ecosistema que permita mapear los distintos usos que se le están dando a los datos de la entidad, y medir si se cumplió con el impacto esperado.
- ✓ Introducir de forma gradual más parámetros de medición impulsados por ciudadanos, que representen las experiencias reales del usuario para pasar de una perspectiva institucional a una ciudadana sobre los resultados e impactos de los datos abiertos.
- ✓ Adoptar un componente de caso de éxito con datos que permita apoyar una cultura de monitoreo y gestión del desempeño basado en datos. Dicho componente debe incluir un perfil de proyecto, alineación estratégica y normativa, análisis costo beneficio, puesta en marcha de los servicios y evaluación y gestión de riesgos.
- ✓ Adelantar, financiar o participar en estudios que permitan cuantificar el valor económico de los datos abiertos en el sector al que pertenece la entidad.
- ✓ Establecer alianzas con entidades públicas y privadas, así como organizaciones de la sociedad civil, academia, ONG para construir proyectos de explotación de datos que den solución a una problemática pública de alto impacto.

CONTROL DE CAMBIOS

Versión	Fecha	Numerales	Descripción de la modificación
00	2021-05-07	Todos	Se crea el Documento metodológico

CONTROL DE APROBACIÓN

Elaboró	Revisó	Aprobó
Nombre - Cargo / Rol: Gabriel Bustos Quiroga – Contratista de la Oficina de Tecnologías y Sistemas de Información	Nombre - Cargo / Rol: Gabriel Bustos Quiroga – Contratista de la Oficina de Tecnologías y Sistemas de Información	Nombre - Cargo / Rol: Elvia Consuelo Castañeda Camargo – Jefe de la Oficina de Tecnologías y Sistemas de Información