

Plan de Datos Abiertos

uts

Unidades
Tecnológicas
de Santander

iLo hacemos posible!

PRESENTACIÓN

Las Unidades Tecnológicas de Santander, en concordancia con los lineamientos del Ministerio de las TIC en favor del mejoramiento en la atención al ciudadano, enfrenta grandes desafíos en la gestión institucional.

Año tras año el mundo evoluciona y exige del Estado un mayor compromiso con la mejora de sus servicios. La apertura de datos institucionales contribuye a la creación del Estado Abierto el cual representa una oportunidad para la ciudadanía por el manejo del *big data*.

El análisis de datos se ha convertido en una ciencia, quienes se dedican a ello examinan los conjuntos de datos con el propósito de sacar conclusiones sobre la información para poder tomar decisiones, o simplemente ampliar los conocimientos sobre diversos temas y presentar las cifras que aporten a los ciudadanos en los diferentes escenarios de la vida.

Las UTS propenden por el mantenimiento de la interacción Estado – Ciudadano, una relación donde se le garantice a la comunidad en general la satisfacción de sus necesidades, el respeto y trato digno al ciudadano, la efectividad de sus derechos, el cumplimiento de sus obligaciones y la facilitación de la gestión de los servidores públicos, por formula el presente Plan de Datos Abiertos.

Las formas tradicionales de Gobierno se han ido transformando hacia el concepto de Estado Abierto en donde la sociedad tiene un rol más activo, y en el que un problema de interés general ya no es solo asunto de los funcionarios públicos sino de la ciudadanía, quien además de aportar nuevos diagnósticos puede solucionarlos de manera colaborativa.

En este nuevo modelo de interacción ciudadano-Estado juegan un rol indispensable la disponibilidad y el acceso a la información pública como facilitadores de la colaboración, la participación y el control social. En Colombia, el acceso a la información pública es un derecho consagrado en la Constitución de 1991 y los datos abiertos son una de las maneras más efectivas de hacer realidad este derecho. Desde el gobierno colombiano se promueve que la información creada, transformada y usada por la administración pública pertenezca a la sociedad; es decir, que el contenido pueda ser consultado, usado y compartido por todos los colombianos, y se encuentre siempre a su disposición, con el fin de que se creen productos o servicios de valor para dar solución a problemáticas públicas.

El presente documento, se enfoca en definir las acciones que las UTS deben abordar para la apertura, estructuración, comunicación y monitoreo de datos abiertos.

1. Plan de Datos Abiertos

1.1. Conceptos clave

- **Accesibilidad:** Garantía de acceso al usuario que lo requiera.
- **Actualidad:** Vigencia de la información.
- **Anonimizar:** Proceso para remover datos personales de una base de datos, buscando la publicación segura de datos para reutilizarlos.
- **Base de datos:** Conjunto de datos pertenecientes a un mismo contexto y almacenados sistemáticamente para su posterior uso.
- **Big data:** es un término en desarrollo que describe un gran volumen de datos estructurados, semiestructurados y no estructurados que tienen el potencial de ser extraídos para obtener información y usarse en proyectos de aprendizaje automático y otras aplicaciones de análisis avanzado.
- **Compleitud:** Información plenamente diligenciada.
- **Comprensibilidad:** Entendimiento e interpretación adecuada de la información por parte de un usuario.
- **Confidencialidad:** Acceso a los datos única y exclusivamente por personas u organizaciones autorizadas, con la intención para proteger adecuadamente la información reservada y clasificada.
- **Conformidad:** Cumplimiento de lineamientos y estándares vigentes.
- **Conjunto de datos:** Unidad mínima de información sujeta a carga, publicación, transformación y descarga en plataformas institucionales.
- **Consistencia:** Datos coherentes y libres de contradicción.
- **Credibilidad:** Información veraz y confiable para los usuarios.
- **Dato:** Descripción de hechos, situaciones, sucesos o valores, representados mediante símbolos físicos o electrónicos.
- **Datos Abiertos:** 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 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.

- **Disponibilidad:** Garantía hacia los usuarios autorizados para tener acceso a la información en el lugar, momento y forma que sea requerida.
- **Eficiencia:** Capacidad para realizar análisis y descargas de los datos con unos niveles de desempeño y tiempos esperados.
- **Entidad:** Institución u organización con la capacidad y/o facultad de definir inventarios de y conjuntos de datos e información a publicar.
- **Formato Libre:** Formato de archivo que se puede crear y manipular mediante cualquier software libre, sin restricciones legales.
- **Gobierno Abierto:** Doctrina política que sostiene que los temas de Gobierno y administración pública deben ser abiertos a todos los niveles posibles en cuanto a transparencia.
- **Información Pública:** Agrupación ordenada de datos públicos, que permite otorgarle a los datos una utilidad y uso en determinado contexto, y que se genera a partir del desarrollo de actividades para el funcionamiento del Estado, es decir de los registros periódicos de las actividades misionales de las entidades, o como consecuencia del ejercicio de funciones de rutina en el Estado.
- **Exactitud:** Datos diligenciados correctamente.
- **Portabilidad:** Característica que garantiza que cualquier conjunto de datos esté representado en un Formato sin restricciones para la reutilización de este.
- **Recuperabilidad:** Atributos que permiten mantener y preservar un nivel específico de operaciones y de calidad.
- **Relevancia:** Utilidad para los usuarios.
- **Trazabilidad:** Histórico del conjunto de datos disponible, fechas de creación, publicación y actualizaciones.

1.2. Marco normativo

La Ley Estatutaria 1712 de 2014 de Transparencia y Acceso a la Información Pública, ha impulsado y exigido la apertura de datos por parte de todas las entidades públicas; a esta norma se suma a la normatividad asociada a la Política de Gobierno Digital y los documentos Conpes que buscan impulsar la explotación

de datos como activos para generar valor social y económico (Conpes 3920 de 2018). Este marco normativo está desarrollado a través de guías y lineamientos que orientan a las entidades a implementar sus planes de apertura y uso de datos abiertos; entre estos se destacan, la guía de calidad de datos, la guía de *linked open data* y el inventario de herramientas libres para analítica.

2. Objetivos

2.1. Objetivo general

Propender por una gestión institucional moderna, eficiente y transparente como instrumento para optimizar el desarrollo de las políticas públicas con el fin de proporcionar una eficiente y eficaz apertura, acceso, uso, apropiación y aprovechamiento de conjuntos de datos abiertos disponibles, como activos públicos, de calidad, oportunos e innovadores para la generación de valor social, cultural, ambiental, y económico.

2.2. Objetivos específicos

- Organizar internamente la información. Promoviendo la eficiencia en la documentación y clasificación de datos
- Permitir la reutilización de los datos publicados. Garantizando Transparencia por posibilitar el análisis y evaluación de la gestión institucional.
- Generar valor. Obteniendo productos derivados de los datos por parte la ciudadanía en general y de empresas.

3. Qué son los Datos Abiertos

Son todos aquellos datos primarios, es decir que están 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 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 de valor agregado. Es decir, los particulares desarrollan aplicaciones, a partir de los datos abiertos publicados por las entidades estatales y benefician a los ciudadanos o usuarios.

La apertura de datos trata de:

- Identificar los conjuntos de documentos, información y datos públicos, gestionados por las entidades públicas.
- Implementar las soluciones técnicas necesarias para ponerlos a disposición de los sistemas de información de terceras partes, como, por ejemplo: aplicaciones para ciudadanos que consumen dicha información.
- Los datos del Gobierno son considerados abiertos si son publicados siguiendo los siguientes principios o características:

Completos: Los datos públicos no deben estar sujetos a privacidad u otras limitaciones. Además, deben estar electrónicamente almacenados.

Primarios: Significa que debe haber una disponibilidad de la fuente primaria, sin procesamientos y sin formas agregadas.

Oportunos: Ello para preservar su valor.

Accesibles: La disponibilidad debe ser lo más amplia posible para los usuarios y para propósitos diversos.

Procesables: Deben estar razonablemente estructurados para permitir su automatización por diversas herramientas.

Acceso indiscriminado: Implica su disponibilidad para cualquier usuario, sin necesidad de su registro.

No-propietarios: Deben estar disponibles en un formato donde nadie deba tener la exclusividad de su control.

Libres de licencias: No deben estar sujetos a *copyright*, patentes, o regulaciones secretas. Una razonable privacidad, seguridad y restricciones puede ser aplicada por el gobierno u otros estamentos.

3.1. Datos Abiertos y Lenguaje Común de Intercambio de Información

El Ministerio de TIC en colaboración con otras Entidades del Estado, han definido un Lenguaje Común de Intercambio de Información, una arquitectura de información y un vocabulario extensible para el intercambio de información entre entidades puede aplicarse a:

- a) La modelación de nuevos conjuntos de datos, en formato GEL-XML.
- b) Su transformación a formatos de datos abiertos.
- c) Su transferencia a plataformas de datos abiertos.
- d) Su vinculación y publicación automatizadas en la Web de los Datos.

Por tanto, cuando se plantea que se deben abrir los datos de la forma más eficaz significa que se deben disponer en todos los formatos estructurados disponibles, simultáneamente. Es decir, se recomienda la publicación de datos en formato GEL-XML y, simultáneamente en formato lo más estándar posible, siendo actualmente ese estándar el RDF. Y dada la gran cantidad de datos ya expresado en el lenguaje común de intercambio de Información (GEL-XML) es una fuente excelente para obtener datos RDF (RDF_XML) mediante transformación de formatos.

Este lineamiento del Ministerio de Tecnologías de la Información y las Comunicaciones requiere la implementación de nuevas herramientas que impulsen el uso del Lenguaje Común de Intercambio para modelar y expresar los datos que se intercambian entre las entidades, algunos de los cuales tienen por destino ser publicados en la Web datos.gov.co. Estas herramientas permitirían la normalización del uso del Lenguaje Común de Intercambio de Información por parte de las entidades.

Es decir, para avanzar en la iniciativa de los datos abiertos, desde las Entidades del Estado debe fortalecerse el uso de Lenguaje Común de Intercambio de Información como un instrumento que permitirá homologar y facilitar la apertura de datos en tanto se va dando la evolución de la Web de los Datos a la Web Semántica.

3.2. Licencia de Datos Abierto para las Unidades Tecnológicas de Santander

Por otro lado, la definición del tipo de licencia bajo el cual se usarán estos datos abiertos se convierte en una de las principales actividades. 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.

Para el licenciamiento de los datos abiertos una alternativa es usar licencias abiertas tipo Creative Commons (CC), que proporcionan un conjunto estándar de términos y condiciones que los creadores y otros titulares de derechos, pueden utilizar para compartir obras originales de su autoría y cualquier otro material sujeto a derechos de autor y a derechos conexos que se especifican en la licencia pública.

La adopción de 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.

3.3. Responsable de los Datos Abiertos

La apertura y administración de los datos abiertos es responsabilidad inicialmente de las Unidades Tecnológicas de Santander, quien a través de cada una de sus dependencias y en mediación de la coordinación de Comunicaciones e Imagen Institucional, se encarga del proceso de apertura, estructuración, comunicación y monitoreo de datos abiertos. A través del portal del estado de datos abiertos www.datos.gov.co

Este plan se actualizará a medida que se vayan identificando nuevas necesidades de información requerida por los ciudadanos, se realizará bimensualmente un seguimiento a este plan a fin de poder ver como se está llevando a cabo el cumplimiento del mismo, para perfeccionar objetivos, plantear nuevos datos, perfeccionar los que se han subido, actualizar los que necesitan dicha actualización entre otras.

El marco de relevancia indica que tan importante es poner en marcha para la Institución el proceso de adecuar los datos abiertos a fin de poder ser publicados en el portal de Datos Abiertos <https://www.datos.gov.co>, se les harán promoción por medio digitales y de comunicación masiva a fin de lograr que la ciudadanía conozca y de uso a estos datos.

También se propenderá por hacer ejercicios de innovación donde se involucren estos datos abiertos generados por la administración a fin de concientizar a los desarrolladores, investigadores y a la sociedad en general de la importancia y calidad de los datos que genera a las UTS.

3.4. Beneficios del Plan de Datos Abiertos

Para las Unidades Tecnológicas de Santander, la reutilización de la información contribuye a incrementar su transparencia y la disminución de tiempo de atención requerido por el usuario.

Para los ciudadanos, ofrece la posibilidad de mejorar su información y conocimiento acerca de las acciones institucionales, haciéndola más transparente y cercana al usuario.

Para las empresas, constituye una base adicional para servicios innovadores y la creación de nuevos servicios de valor agregado.

4. Marco de Gobernanza

Desde el nacimiento de la iniciativa de datos abiertos en 2011, el Ministerio TIC ha establecido un marco de gobernanza para garantizar la sostenibilidad y promover el impacto de esta. Este marco está soportado en cinco componentes claves: 1) marco normativo, 2) actores, 3) herramientas e incentivos y 4) monitoreo:

4.1. Marco Normativo.

Desde la Ley Estatutaria 1712 de 2014 de Transparencia y Acceso a la Información Pública, se impulsa y exige la apertura de datos por parte de todas las entidades públicas; a esta ley estatutaria se suma la normatividad asociada a la Política de Gobierno Digital y los documentos Conpes relacionados a esta política que buscan impulsar la explotación de datos como activos para generar valor social y económico (Conpes 3920 de 2018). Este marco normativo está desarrollado a través de guías y lineamientos que orientan a las entidades a implementar sus planes de apertura y uso de datos abiertos; entre estos se destacan, la guía de calidad de datos, la guía de *linked open data* y el inventario de herramientas libres para analítica.

4.2. Actores.

Motor de esta iniciativa, en la cual periodistas, estudiantes, emprendedores, sociedad civil y gobierno convergen en la consecución de un mismo objetivo: más y mejor información y usos de esta. Los actores gubernamentales de este ecosistema están encabezados por el Ministerio TIC, a través de la Dirección de Gobierno Digital, como responsable del diseño y la promoción de la iniciativa, y acompañado por otros líderes de política como la Secretaría de Transparencia, el Departamento Administrativo de la Función Pública - DAFP, el Departamento Nacional de Estadística - DANE y el Departamento Nacional de Planeación - DNP.

4.3. Herramientas e Incentivos.

El Ministerio TIC ha dispuesto el Portal de Datos del Estado Colombiano <http://www.datos.gov.co/>, como el espacio digital para la disposición de la información pública estructurada del país y el punto de encuentro del ecosistema para su uso y aprovechamiento. Éste es operado y administrado por la Dirección de Gobierno Digital y reúne las herramientas que se proveen a los ciudadanos y entidades para publicar, acceder y usar los datos abiertos de Gobierno.

Durante la vigencia 2020 los datos existentes en *datos.gov.co* fueron sometidos a una detallada revisión para su depuración en procura del mejoramiento de la

calidad de los datos abiertos disponibles. El MinTic creó la *Guía de Estándares de Calidad e Interoperabilidad de los Datos Abiertos del Gobierno Colombiano*, la cual detalla la creación de Datos abiertos para la transparencia y el control social, para empoderar al ciudadano, para mejorar o crear productos y modelos de negocio, datos como una puerta abierta a la innovación, para mejorar la eficiencia y eficacia del Estado, y para medir el impacto de las políticas.

4.4. Monitoreo

Ministerio TIC cuenta con algunas alternativas para monitorear la apertura y el uso de los datos abiertos tales como:

FURAG: el Formulario Único Reporte de Avances de la Gestión, mide anualmente a las entidades del orden nacional y territorial el avance en las políticas de desarrollo administrativo, incluidos los indicadores de proceso y resultado relacionados con la apertura y uso de datos abiertos en el marco de la Política de Gobierno Digital y Acceso a la Información Pública. Los resultados de la medición están disponibles en:

Sello de excelencia de Gobierno Digital: es un modelo de madurez que garantiza y certifica la alta calidad de los productos y servicios digitales de las entidades públicas del Estado Colombiano.

El Sello fue creado a través del Decreto 2573 de 2014 y adoptado en la Resolución 2405 de 2016. Se encarga de certificar la alta calidad de los siguientes productos en 5 Categoría:

- Datos abiertos
- Ejercicios de participación por medios digitales
- Trámites y Servicios
- Capacidades de Gestión TI
- Territorios y Ciudades Inteligentes

Evaluación ITA: Índice de Transparencia y Acceso a la Información, realizado por la Procuraduría General de la Republica. El cual entre sus puntos califica la utilización de Datos Abiertos en los Portales Institucionales dentro del Botón de transparencia y acceso a la Información.

5. Ciclo de los Datos Abiertos

5.1. Fase 1 - Establecer el plan de apertura.

Las dependencias de las Unidades Tecnológicas de Santander identifican la información relevante y de alto impacto de la institución para la Ciudadanía.

- Información que pueda generar valor para terceros.
- Ser fuente primaria de la información.
- Información que sea solicitada o necesitada por los ciudadanos.
- Calidad de la Información.
- Esfuerzo requerido para publicar. Validación de Información estadística publicada en el portal Web de la entidad. Información rotulada como pública por el Índice de Información Clasificada y Reservada de la Institución.

Una vez identificados los datos abiertos de las Unidades Tecnológicas de Santander, se debe 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).

5.2. 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 eliminar la información sensible que afecta la privacidad de personas u organizaciones puesto que anonimizar los datos es obligatorio antes de su publicación.

5.3. Priorizar

Los datos de valor son importantes para que los esfuerzos de gobernanza, apertura y publicación creen impactos positivos en la ciudadanía y permitan solucionar problemáticas públicas.

La priorización debe realizarse tomando en cuenta los siguientes criterios:

- Ofrecer datos que permiten dar cumplimiento a metas institucionales de transparencia, participación, control social y colaboración.
- Verificar si los datos contribuyen a un propósito de política pública específico y que hacen parte de las categorías establecidas por las mejores prácticas a nivel nacional e internacional en la materia y aquellos solicitados directamente por otras entidades públicas.
- Favorecen al cumplimiento de las metas trazadas en el Plan Nacional de Desarrollo y los planes departamentales y municipales.
- Aportan al cumplimiento de los Objetivos de Desarrollo Sostenible.

- Son datos que han sido solicitados por los ciudadanos en ejercicios virtuales o presenciales de consulta pública.

6. Estructurar y publicar los datos

Una vez se tiene el inventario de datos abiertos priorizados, se hace la preparación para publicar los datos abiertos, la cual consiste en alistar los datos para que éstos puedan ser procesados y reutilizados por terceros.

6.1 Documentar

Los metadatos 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.

El Portal <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.

6.2 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 usados por los reutilizadores. No deben utilizarse formatos cerrados o no editables como por ejemplo información en formato PDF.

6.3 Publicar

Acá se pone a disposición de los usuarios el conjunto de datos abiertos a través del cargue de datos en el portal www.datos.gov.co, este es el sitio dispuesto por el Ministerio TIC, donde se deben publicar los datos abiertos de interés nacional.

Carga manual de datos: Este proceso se realiza a través de la herramienta de carga dispuesta en www.datos.gov.co, con el uso del usuario y contraseña proporcionado a la Institución.

7. Comunicar y promover el uso de los datos

La Institución debe poner a disposición de sus usuarios sus datos y establecer mecanismos que permitan identificar y difundir los productos, servicios, artículos periodísticos, investigaciones académicas y en general cualquier producto que use con ellos.

Es por ello que los datos abiertos institucionales deben estar publicados en la sección de transparencia del Portal Institucional además del Portal Datos del Estado Colombiano www.datos.gov.co, para promover la transparencia, el acceso a la información pública, entre los funcionarios, contratistas, partes interesadas y ciudadanía en general dando una mayor apertura y orientando que los datos sean lo más ricos y completos posibles.

8. Monitorear la calidad y el uso

Cuando se garantiza 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.

Las Unidades Tecnológicas de Santander deben 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.

Debe monitorearse la calidad de los datos abiertos publicados, es posible utilizar algunos indicadores de tipo cuantitativo cómo:

- Número de solicitudes de apertura de datos.
- Número de dependencias u oficinas de la Institución que abren datos.
- 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 frecuencia pero que el ciudadano decide no descargar, para revisar si el conjunto tiene temas de calidad por mejorar o comentarios ciudadanos solicitando mejoras puntuales.

Para mapear el impacto que está teniendo el uso y aprovechamiento de datos abiertos, existe una sección de usos de datos abiertos, dispuesta en www.datos.gov.co que permite 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, pueden usarse algunos indicadores de tipo cuantitativo que permiten determinar los productos desarrollados a partir de sus datos abiertos tales como:

- Número de soluciones web o móviles generadas a partir de datos abiertos.
- Número de proyectos que hicieron uso de los conjuntos de datos abiertos.
- Número de proyectos de emprendimiento o nuevos negocios generados a partir del uso de datos abiertos.
- Número de publicaciones que hicieron uso de los conjuntos de datos abiertos (papers, artículos, noticias, libros, etc.).