**Estudio sobre minería y visualización de datos abiertos del gobierno de Colombia**Anteproyecto
Línea de investigación del grupo FICB-PG:
Grupo de Investigación FICB-PG **Yeimmy Esperanza Lee Olaya (1610020260)**Asesor temático: ING MSC Laura Juliana Cortés Rico

Supervisor metodológico: ING MSC MBA PHD.C Luis Gabriel Moreno Sandoval

*Enero 2018*

**\_\_\_\_\_\_\_\_\_\_\_\_**

**RESUMEN**

En este estudio se realizará una revisión sistemática de los conjuntos de datos publicados en el portal de datos abiertos de Colombia, definido en términos de la calidad de la información que está siendo compartida por las entidades del gobierno y de las herramientas de minería y visualización que permitan promover y facilitar su uso siendo un instrumento que apoye el cumplimiento de los tres pilares del gobierno abierto: Transparencia, Participación y Colaboración.

**Palabras clave:** gobierno abierto, datos abiertos, minería de datos, visualización de datos.

**ABSTRACT**

In this study, a systematic review of the data sets published in the open data portal of Colombia will be carried out, defined in terms of the quality of the information that is being shared by the government entities and the data mining and data visualization tools that allow to promote and facilitate its use as an instrument that supports the fulfillment of the three pillars of open government: Transparency, Participation and Collaboration.

**Key words:** open government, open data, data mining, data visualization.

**\_\_\_\_\_\_\_\_\_\_\_\_**

**INTRODUCCIÓN**

Los datos abiertos se definen como datos digitales, primarios o sin procesar, en formatos legibles por máquinas, publicados por entidades públicas con las características técnicas y jurídicas necesarias para ser usados, reutilizados y redistribuidos por cualquier persona, en cualquier momento y lugar.

Desde el año 2015, el gobierno nacional, por medio del Ministerio de Tecnologías de la Información y las Comunicaciones MinTIC, lanzó el portal [www.datos.gov.co](http://www.datos.gov.co), y con esto, abrió un sinnúmero de posibilidades para que cualquier ciudadano, periodista, estudiante o empresario, de valor a esos datos y genere productos y servicios derivados a partir de estos.

Este proyecto consiste en aplicar técnicas de minería y visualización de datos sobre los datos abiertos del gobierno de Colombia con el propósito de incentivar el uso y la difusión de los mismos. Para esto, se debe hacer una revisión sistemática a una muestra representativa de los conjuntos de datos publicados, definiendo unas métricas para evaluar la calidad y un proceso automático que permita diagnosticar y aplicar las técnicas de visualización apropiadas para cada conjunto.

1. **PLANTEAMIENTO DEL TEMA Y JUSTIFICACIÓN DEL ESTUDIO**

Los datos abiertos promueven la transparencia, la colaboración y la participación, son el mecanismo que brinda el gobierno para que ciudadanos y empresas generen valor a partir de la información producida por las entidades del Estado. La apertura de datos fortalece la democracia y promueve la eficiencia y efectividad del gobierno (Obama, 2009)

En palabras de Daniel Quintero (Exviceministro de Economía Digital) los datos son el nuevo petróleo, la iniciativa de datos abiertos es la oportunidad para abrir esos datos y entregárselos a los ciudadanos en tiempo real, de modo que los ciudadanos sean los que saquen provecho de esta información: hagan control a los gobiernos, la utilicen para crear empresa y generar riqueza. Estos datos pueden ser utilizados por periodistas para hacer control político, estudiantes que quieran hacer investigaciones, emprendedores que necesiten información del estado y por el mismo sector público para crear políticas obteniendo información de qué están haciendo todas las entidades (Quintero, 2016).

De acuerdo a la Guía de datos abiertos en Colombia (MINTIC, 2016), el círculo virtuoso de la apertura de datos contempla: datos, reúso y valor. Una estrategia de datos abiertos va más allá de la preparación y publicación de los datos por parte de las entidades. También involucra acciones concretas para promover el uso de los datos publicados, realizar seguimiento e identificar el valor agregado que se está generando con su utilización. En este caso se va a centrar la atención en la utilización de los datos publicados. A la fecha, en el portal de datos abiertos de Colombia hay más de 4700 conjuntos de datos de más de 980 instituciones del gobierno (Colombia D. A., 2017).

Los datos abiertos deben ser expuestos en formatos abiertos que puedan ser leídos por humanos y por máquinas: XML, RDF, CSV (W3C, 2009). Sin embargo, ver muchas filas y columnas en un archivo separado por comas, aunque es muy útil para que una máquina lo lea y procese, para un ser humano no es una actividad amigable ni productiva. Para encontrar patrones, producir conocimiento y tomar decisiones, se necesita que la información tenga un procesamiento y muestre unos resultados sintetizados al lector.

Según Stephen Few, fundador de Perceptual Edge, firma de consultoría sobre visualización de datos y autor de varios libros relacionados, esta es la era de ‘muchos datos, pero poca información’[[1]](#footnote-1). La mayoría de la gente se encuentra sepultada en montañas de datos sin saber cómo usarlos. Quienes saben darles sentido a los datos, son conscientes de la gran oportunidad que se tiene de tomar mejores decisiones y de progresar como nunca antes. Se debe aprender a aprovechar la corriente constante de datos, de manera que se pueda usar lo aprendido para mejorar el mundo. Los datos no califican como información si no informan. Para usar los datos de forma efectiva, primero se les debe dar sentido. (Few, 2012)

Para darles sentido, existe la minería de datos, la cual es una disciplina que desarrolla métodos computacionales para extraer información valiosa de datos en bruto, por medio de algoritmos de análisis automático (Keim Daniel & Mansmann, 2010). Para que la minería de datos sea efectiva, es importante incluir a las personas en el proceso de exploración de datos y combinar la flexibilidad, creatividad y conocimiento general de las personas con la enorme capacidad de almacenamiento y poder computacional de los ordenadores actuales.

Otra disciplina valiosa para este fin es la visualización de datos, que se utiliza para aprovechar la capacidad humana de extraer patrones a partir de imágenes y también para ayudar al usuario final a comprender los resultados del análisis. La idea básica es presentar los datos en alguna forma visual, permitiendo a las personas tener una visión completa, sacar conclusiones e interactuar directamente con ellos (D.A. Keim, 2002).

Los datos ya están disponibles, ahora se deben convertir en información útil para sacarles el mayor provecho. Teniendo en cuenta que la cantidad de conjuntos de datos publicados es masiva y crece más cada día, es momento de pensar si es posible automatizar hasta cierto punto el análisis básico de los datos abiertos, sin querer desplazar en ningún momento al científico de datos, el cual cumple un rol fundamental en el proceso de minería de datos, pero si poder llegar a un escenario donde se aproxime la comprensión y el uso de la información del gobierno a un ciudadano del común, sin necesidad de que sea un científico de datos. Otra ventaja de automatizar el análisis, es poder proveer realimentación continua a las propias entidades que publican los conjuntos de datos, lo cual ayudaría a mejorar la calidad de los mismos.

1. **PREGUNTA DE INVESTIGACIÓN**

¿Cómo se podría automatizar el proceso de análisis básico de los conjuntos de datos publicados en el portal de datos abiertos de Colombia para entregar los resultados en formatos gráficos y que personas no expertas en analítica de datos puedan usarlo?

* 1. **Variables**
* Automatización de analítica básica de datos abiertos
* Visualización de datos
1. **OBJETIVOS**
	1. **Objetivo General**

Realizar un estudio sobre visualización automatizada de datos de los conjuntos publicados en el portal Datos Abiertos de Colombia, definido en términos de la información que está siendo compartida por las entidades del gobierno y de las herramientas de minería y visualización que promuevan y faciliten su uso.

* 1. **Objetivos Específicos**
* Analizar herramientas existentes para visualización de datos, generando un reporte comparativo de características y costos.
* Definir métricas que permitan evaluar la calidad de los conjuntos de datos que se están compartiendo por parte de las entidades del estado.
* Diseñar un proceso automático que permita generar la descriptiva básica de las columnas, identificar valores inconsistentes, incompletos y con errores a cualquier conjunto de datos publicado en el portal.
* Hacer una revisión sistemática sobre una muestra de los conjuntos de datos publicados en el portal, empleando las métricas definidas y el proceso automático diseñado.
* Establecer y aplicar las técnicas de visualización para cada conjunto de datos de acuerdo a la revisión.
1. **ESTRATEGIA METODOLÓGICA**

En cuanto al alcance de la investigación, ésta se determina como una investigación de enfoque mixto con preponderancia cuantitativa, básicamente por las siguientes razones:

Un enfoque mixto busca lograr una perspectiva más amplia de la realidad que se va a estudiar, esto se consigue explorando desde diversas posturas y de forma completa, la profundidad y la complejidad del problema a solucionar con el objetivo de la comprensión de la relación entre la selección de las herramientas adecuadas de minería de datos y el incentivar a los ciudadanos para que usen los datos abiertos. Los procesos que se desarrollen deben estar orientados a complementar los resultados analizando datos tanto cualitativos como cuantitativos (Hernandez Sampieri, Fernandez Collado, & Baptista Lucio, 2010).

Se requiere hacer levantamiento de información estructurada para el análisis de los datos cuantitativos, para identificar, por ejemplo: cantidad de personas que utilizan actualmente los conjuntos de datos publicados, cantidad de técnicas de minería existentes, tipos de gráficas según los tipos de datos manejados; lo anterior combinarlo con entrevistas a los distintos interesados (periodistas, estudiantes, servidores públicos, desarrolladores de aplicaciones) sobre la percepción cualitativa que se tiene sobre el uso de dichos conjuntos de datos. Al obtener información numérica y las observaciones cualitativas se enriquece la comprensión de la necesidad y permite ofrecer una respuesta con mayor pertinencia.

La investigación se plantea con preponderancia cuantitativa de corte descriptivo, a partir de las siguientes razones según (Hernandez Sampieri et al., 2010):

* Establece la relación entre dos variables: Automatización de analítica básica de datos abiertos y Visualización de datos.
* El problema identificado que relaciona las dos variables identificadas se expresa en forma de pregunta: “¿Cómo se podría automatizar el proceso de análisis básico de los conjuntos de datos publicados en el portal de datos abiertos de Colombia para entregar los resultados en formatos gráficos y que personas no expertas en analítica de datos puedan usarlo?”.
* El planteamiento implica la realización de pruebas empíricas, es decir, ofrece la factibilidad de observar una realidad de manera objetiva que está determinada por un contexto específico (Conjunto de datos portal de datos abiertos de Colombia) y una población delimitada (interesados en datos abiertos: periodistas, estudiantes, servidores públicos, desarrolladores de aplicaciones); donde la meta es describir las particularidades del contexto para generalizar y replicar en donde persiste una posición neutral y distante del investigador con respecto al problema planteado.

En cuanto al diseño investigativo, el estudio se describe como una investigación no experimental descriptiva por cuanto éste no pretende la manipulación de variables deliberadamente, sino la obtención de detalles en contexto para identificar las características de las técnicas de minería y las características de la visualización de datos, para luego analizar. En este caso las herramientas de minería son basadas en lo que los autores proponen, y para el uso se acudirá a personas que ya hayan tenido contacto con esta plataforma de datos abiertos o con otras; de esta forma las inferencias sobre las relaciones entre estas dos variables se realizan sin intervención o influencia directa, y dichas relaciones se observan y analizan tal como se han dado en su contexto natural.

## Recolección de datos

Revisión literaria: para poder identificar y clasificar las herramientas de minería de datos, e inclusive estudios previos sobre uso de datos abiertos a nivel mundial.

Observación: se debe explorar una muestra de los conjuntos de datos existentes en el portal para poder hacer un análisis y una posible clasificación.

Entrevistas: es muy importante conocer la opinión de los interesados en datos abiertos.

De acuerdo a lo anterior se definen los siguientes pasos como estrategia metodológica:

1. Estado del arte

Validación de otros estudios similares en otros países para poder determinar los parámetros de evaluación de los conjuntos de datos.

2. Análisis de uso

Encuestas sobre el uso de datos abiertos en Colombia y entrevistas a las partes interesadas: periodistas de datos, estudiantes, emprendedores, funcionarios públicos y ciudadanos en general.

3. Análisis de conjuntos de datos

Revisión sistemática de los conjuntos de datos del portal de datos abiertos de Colombia, aplicando herramientas de minería y de visualización de datos.

4. Definición del proceso automático de evaluación

Definición de las diferentes etapas que permitan la evaluación de los conjuntos de datos, apoyado en los datos obtenidos en las etapas previas (paso 1, 2 y 3).

1. **MARCO CONCEPTUAL**

## Datos abiertos

Los datos abiertos son datos que pueden ser utilizados libremente, reutilizados y redistribuidos por cualquier persona (Handbook, s.f.).

La definición de datos abiertos según la ley colombiana de transparencia y acceso a la información dice que son “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” (Congreso de la República, 2014).

Según la Carta Internacional de Datos Abiertos, sus principios son (Open Data Charter, s.f.):

1. Abiertos por Defecto
2. Oportunos y Exhaustivos
3. Accesibles y Utilizables
4. Comparables e Interoperables
5. Para mejorar la Gobernanza y la Participación Ciudadana
6. Para el Desarrollo Incluyente y la Innovación

### Historia

Los datos abiertos son uno de los elementos del gobierno abierto, y aunque la primera vez que se habló de gobierno abierto fue en el espacio político británico a finales de la década de los 70, fue a principios del siglo XXI, que los países empezaron a generar reglamentación e iniciativas para que el gobierno usara las tecnologías de la información y las comunicaciones para mejorar su gestión de cara a la ciudadanía (Ramírez-Alujas, 2011). En el año 2009, el término datos abiertos, empezó a tomar fuerza con el discurso del entonces presidente de los Estados Unidos, Barack Obama, sobre transparencia y gobierno abierto (Obama, 2009). Por ese mismo año, Tim Berners-Lee, mejor conocido como el padre de la web (W3C, s.f.), ofreció una charla por medio de la plataforma TED, donde hablaban de la importancia de la apertura de datos y además de enlazar estos datos para generar más conocimiento: “hablamos de hacer que el mundo funcione mejor haciendo que los datos estén disponibles” (Berners-Lee, 2009).

En la actualidad, según la Open Knowledge International, que desde el 2013 publica el Global Open Data Index, más de 120 países tienen portales de datos abiertos, en el año 2015 Colombia ocupaba el puesto 4 en el mundo y primero en Latinoamérica (OKFN, 2015). Para el año 2016 se agregaron unos indicadores adicionales al índice, ubicando a Colombia en el puesto 12 a nivel mundial, segundo en Latinoamérica, superado únicamente por Brasil que pasó de estar en el puesto 12 al puesto 7 (OKFN, 2016).

### Datos abiertos en Colombia

Desde inicios del milenio, el gobierno colombiano ha establecido reglamentación que busca masificar el uso de las Tecnologías de la Información y con ello aumentar la competitividad del sector productivo, modernizar las instituciones públicas y de gobierno, y socializar el acceso a la información, para esto en el Consejo Nacional de Política Económica y Social CONPES de febrero del 2000, proponen la política denominada “el salto a internet” (CONPES, 2000).

En el año 2008 se generaron los lineamientos generales de la Estrategia de Gobierno en Línea de la República de Colombia, mediante el decreto 1151 (Presidencia, 2008). En el año 2009, el 30 de julio, bajo la ley 1341 el Ministerio de Comunicaciones pasa a ser Ministerio de Tecnologías de la Información y las Comunicaciones (MINTIC, 2014). Sin embargo, hasta el año 2015 se empieza a hablar de apertura de datos, con el lanzamiento del Portal de Datos Abiertos de Colombia www.datos.gov.co (MINTIC, 2015). Hoy en día hay más de 5000 conjuntos de datos publicados por más de 900 entidades.

En la actualidad, el gobierno está impulsando no sólo la publicación sino también el uso de los datos para darle dinamismo al círculo virtuoso de la apertura de datos; por ejemplo, el proyecto ‘Emprende con datos’ creado para fomentar el desarrollo de productos digitales, aplicaciones web y/o móviles, a través de acompañamiento profesional en mentorías para emprendedores TIC, ayudando a resolver problemáticas de interés público y social a partir del uso de Datos abiertos de gobierno. Está dirigido a emprendedores, entidades públicas y pequeñas empresas del país que hagan parte del ecosistema TIC (MINTIC Y FINDETER, 2016).

Un caso de éxito con los datos abiertos, es el municipio Tamalameque, Cesar. En abril del 2016, un reconocido medio de comunicación publicaba una noticia que informaba que en dicho municipio se consumía la peor agua del país (El Tiempo, 2016). Al darle una relevancia nacional, la administración del municipio tomó cartas en el asunto y en noviembre del mismo año se anunció el inicio del proceso de cloración del agua como parte de un plan de acción para mejorar la calidad del agua de la zona (El Pilón, 2016).

El portal de datos abiertos de Colombia está implementado sobre un software propietario diseñado para tal fin, llamado Socrata, el cual es una plataforma Software como servicio SaaS, proporciona publicación y manipulación de datos; ofrece visualizaciones básicas: gráficos de barras, listas y mapas; permite guardar las vistas generadas; y tiene un API de consulta de datos (MINTIC, 2016).

## Visualización en minería de datos

La visualización de datos es un conjunto de métodos que utilizan representaciones visuales para explorar, dar sentido y comunicar datos cuantitativos (Few, 2017). La visualización en minería de datos es la técnica de exploración y análisis de gran volumen de datos con esfuerzos combinados de máquina y cerebro humano. Se utiliza para extraer información útil para la toma de decisiones (Kumar, 2016).

Las ventajas de la visualización de datos son (D. Keim & Ward, 2002):

* Maneja datos muy poco homogéneos y con mucho ruido.
* Es intuitiva y no requiere la comprensión de complejos algoritmos matemáticos o estadísticos.
* Puede proporcionar una visión cualitativa de los datos, permitiendo aislar los fenómenos en los datos para permitir análisis adicionales.

Se puede clasificar en (Daniel A. Keim, Sips, & Ankerst, 2005):

* Visualización previa: Cuando se usa para antes de aplicar el algoritmo de minería de datos. Está dirigido al experto en minería.
* Visualización posterior: Cuando se usa para mostrar los resultados del proceso de minería de datos. Está dirigido al experto de negocio.
* Visualización totalmente integrada: Cuando se usa en todas las etapas, en el procesamiento de datos, aplicación del algoritmo y reporte de resultados.

La visualización de datos en la etapa previa facilita (Galit Shmueli, Peter C. Bruce, 2016):

* Limpieza de datos encontrando valores incorrectos, faltantes, filas duplicadas, columnas con el mismo valor, y similares.
* Generación de variables derivadas.
* Selección de variables para incluir en el análisis.
* Detección de variables redundantes.
* Identificación de variables a numerizar y/o categorizar.
* Definición de muestras para conjuntos de datos demasiado grandes.

### Tipos de gráficos

Algunos tipos de gráficos que se pueden utilizar para visualización de datos son:

Tabla 1. Tipos de gráficas en Visualización de datos

|  |  |
| --- | --- |
| Gráficos básicos | * Gráfico de barras
* Gráfico de líneas
* Gráficos de dispersión
 |
| Diagramas de distribución | * Diagrama de cajas
* Histogramas
 |
| Visualización Multidimensional | * Mapas de calor
* Gráficos de dispersión con códigos de color
* Gráficos de barras con múltiples páneles
* Matriz de gráficos de dispersión
* Gráfico de coordenadas paralelas
 |
| Visualizaciones especializadas | * Red de datos
* Jerarquías con mapa de árbol
* Gráficos de mapas
 |

Fuente: Resumen de tipos de gráficos propuestos en (Galit Shmueli, Peter C. Bruce, 2016)

## Proyectos de referencia

### Estudio sobre el uso de visualizaciones para datos de gobierno abierto (Graves & Hendler, 2014)

En el Instituto Politécnico Rensselaer de Nueva York, realizaron un análisis donde explican que tener los datos disponibles no es suficiente para que la gente los use, entre otras cosas por falta de conocimiento o destrezas técnicas de los interesados. Ellos proponen la visualización como medio para explorar y compartir datos. Identificaron cinco posibles sujetos interesados: publicador de datos del gobierno, consumidor de datos del gobierno, investigador / periodista, programadores y ciudadano del común.

El primer paso, por medio de encuestas, fue preguntarles a los diferentes interesados si sería un artefacto valioso para ellos. Seguido de esto realizaron entrevistas para entender el contexto en el cual los datos serían usados. También analizaron los programas disponibles para visualización de datos. Finalmente, determinaron desarrollar un prototipo de una herramienta específica para visualización de datos abiertos, el prototipo se llama OpenDataVis y puede ser descargada de forma libre de sourceforce.net.

### ¿Qué se está haciendo con los datos públicos abiertos? Un análisis exploratorio de los usos públicos de los datos abiertos en la ciudad de Nueva York (Okamoto, 2016)

Karen Okamoto del John Jay College en junio del 2016 publicó su investigación sobre el uso de los datos abiertos de la ciudad de Nueva York. Según cuenta en su artículo desde el año 2012 se aprobó la legislación sobre apertura de datos en la ciudad de Nueva York y en 2016 habían más de 1300 conjuntos de datos publicados. En el estudio buscó y analizó páginas web, aplicaciones móviles y visualizaciones que usan datos abiertos. A lo largo del artículo detalla los hallazgos y una serie de estadísticas con los criterios que ella utilizó para la evaluación.

En cuanto a visualización encontró uso de mapas y de infografías. El estudio contempló 77 ítems en total: mapas (48%), aplicaciones web (19%), infografías (9%) y aplicaciones móviles (5%), por mencionar algunos formatos. La conclusión es que la apertura de datos inspira un amplio rango de usos creativos y propone como trabajos futuros medir el impacto de dichos productos. El propósito del estudio era mostrar una captura actualizada de los usos de los datos abiertos, además ampliar la literatura sobre el tema para servir de guía en estudios sobre otras regiones.

* 1. **Marco regulatorio de datos abiertos**

En Colombia la apertura de datos tiene un soporte jurídico para garantizar el derecho a la protección de datos personales y la intimidad, y a su vez, ordenar el deber del Estado de divulgar información pública. Se presenta un cuadro con la normatividad que regula los datos abiertos.

|  |  |
| --- | --- |
| Norma | Descripción  |
| Constitución Política del Colombia (Colombia, 1991) | * Artículo 2. Fines esenciales del estado
* Artículo 15. Derecho a la intimidad
* Artículo 20. Libertad de expresión
* Artículo 27. Libertad de investigación
* Artículo 61. Propiedad intelectual
* Artículo 74. Acceso a documentos públicos
 |
| Ley 1266 de 2008 (Congreso de la República, 2008) | "Por la cual se dictan las disposiciones generales del habeas data y se regula el manejo de la información contenida en bases de datos personales, en especial la financiera, crediticia, comercial, de servicios y la proveniente de terceros países y se dictan otras disposiciones." |
| Ley 341 de 2009 (Congreso de Colombia, 2009) | "Por la cual se definen Principios y conceptos sobre la sociedad de la información y la organización de las Tecnologías de la Información y las Comunicaciones -TIC-, se crea la Agencia Nacional del Espectro y se dictan otras disposiciones”* Artículo 2. Principio 8. Masificación del Gobierno en línea.
 |
| Ley 1450 de 2009 (Congreso de Colombia, 2001) | "Por el cual se expide el Plan Nacional de Desarrollo 2010-2014”* Artículo 230. Gobierno en línea como estratégica de buen gobierno.
 |
| Ley 1437 de 2011 (Congreso de la República, 2011) | “Por la cual se expide el Código de Procedimiento Administrativo y de lo Contencioso Administrativo”* Artículo 8. Deber de información al público.
 |
| Ley 1564 de 2012 (Congreso de la República, 2012) | “Por medio de la cual se expide el Código General del Proceso y se dictan otras disposiciones”* Artículo 103. Uso de las tecnologías de la información y de las comunicaciones
 |
| Ley 1581 de 2012(Congreso de Colombia, 2012) | “Por la cual se dictan disposiciones generales para la protección de datos personales” |
| Ley 1712 de 2014(Congreso de la República, 2014) | "Por medio de la cual se crea la Ley de Transparencia y del Derecho de Acceso a la Información Pública Nacional y se dictan otras disposiciones" |
| Decreto 2573 de 2014(Presidente de la República de Colombia, 2014) | “Por el cual se establecen los lineamientos generales de la Estrategia de Gobierno en línea, se reglamenta parcialmente la Ley 1341 de 2009 y se dictan otras disposiciones.”* Artículo 4. Principios y fundamentos de la Estrategia de Gobierno en línea. Apertura y reutilización de datos públicos.
 |
| Decreto 1074 de 2015(Presidente de la República de Colombia, 2015) | “Por medio del cual se expide el Decreto Único Reglamentario del Sector Comercio, Industria y Turismo”* Capítulo 25. Reglamenta parcialmente ley 1581 de 2012
* Capítulo 26. Registro nacional de bases de datos
* Capítulo 27. Contenido mínimo de las historias crediticias
 |
| Decreto 1081 de 2015(Presidente de la República de Colombia, 2015) | “Por medio del cual se expide el Decreto Reglamentario Único del Sector Presidencia de laRepública”* Titulo 1. Disposiciones generales en materia de transparencia y del derecho de acceso a la información pública nacional
 |
| Ley 1753 de 2015(Congreso de la República de Colombia, 2015) | “Por la cual se expide el Plan Nacional de Desarrollo 2014-2018”* Artículo 45. Estándares, modelos y lineamientos de tecnologías de la información y las comunicaciones para los servicios al ciudadano.
 |
| Ley 1755 de 2015(Congreso de Colombia, 2015) | “Por medio de la cual se regula el Derecho Fundamental de Petición y se sustituye un título del Código de Procedimiento Administrativo y de lo Contencioso Administrativo”* Artículo 24. Informaciones y documentos reservados.
 |
| Resolución 3564 de 2015(MinTIC, 2015) | “Por la cual se reglamentan aspectos relacionados con la Ley de Transparencia y Acceso a la Información Pública”* Anexo 1. Estándares para publicación y divulgación de información.
 |

* + 1. **Habeas Data**

Es el “derecho constitucional que tienen todas las personas a conocer, actualizar y rectificar las informaciones que se hayan recogido sobre ellas en bancos de datos” (Congreso de la República, 2008). En Colombia este derecho está fundamentado en el artículo 15 de la Constitución Política de Colombia. Posteriormente fue desarrollado en las leyes 1266 de 2008 y 1581 de 2012. En la primera se refiere únicamente a información financiera, la segunda es de carácter general y trata sobre el tratamiento cualquier información de un ciudadano.

* + 1. **Tipos de dato**

De acuerdo a la ley 1266 de 2008 (Congreso de la República, 2008).

* Dato Personal

Es cualquier pieza de información vinculada a una o varias personas determinadas o determinables o que pueda asociarse con una persona natural o jurídica.

* Dato Público

Son públicos, entre otros, los datos contenidos en documentos públicos, sentencias judiciales debidamente ejecutoriadas que no estén sometidos a reserva y los relativos al estado civil de las personas.

* Dato Semiprivado

Es el dato que no tiene naturaleza ìntima, reservada, ni pública y cuyo conocimiento o divulgación puede interesar no sólo a su titular sino a cierto sector o grupo de personas o a a la sociedad en general, como el dato financiero y crediticio de actividad comercial.

* Dato Privado

Es el dato que por su naturaleza íntima o reservada sólo es relavante para el titular.

* + 1. **Responsables de la información**

De acuerdo a la ley 1712 de 2014 (Congreso de la República, 2014). Un sujeto obligado se refiere a cualquier persona natural o jurídica, pública o privada incluida en el siguiente listado:

* + Toda entidad pública, incluyendo las pertenecientes a todas las Ramas del Poder Público, en todos los niveles de la estructura estatal, central o descentralizada por servicios o territorialmente, en los órdenes nacional, departamental, municipal y distrital.
	+ Los órganos, organismos y entidades estatales independientes o autónomos y de control.
	+ Las personas naturales y jurídicas, públicas o privadas, que presten función pública, que presten servicios públicos respecto de la información directamente relacionada con la prestación del servicio público.
	+ Cualquier persona natural, jurídica o dependencia de persona jurídica que desempeñe función pública o de autoridad pública, respecto de la información directamente relacionada con el desempeño de su función.
	+ Los partidos o movimientos políticos y los grupos significativos de ciudadanos.
	+ Las entidades que administren instituciones parafiscales, fondos o recursos de naturaleza u origen público.
	+ Las personas naturales o jurídicas que reciban o intermedien fondos o beneficios públicos territoriales y nacionales y no cumplan ninguno de los otros requisitos para ser considerados sujetos obligados, solo deberán cumplir con la presente ley respecto de aquella información que se produzca en relación con fondos públicos que reciban o intermedien.

No serán sujetos obligados aquellas personas naturales o jurídicas de carácter privado que sean usuarios de información pública.

* + 1. **Tipos de información**

De acuerdo a la ley 1712 de 2014. La información se refiere a un conjunto organizado de datos contenido en cualquier documento que los sujetos obligados generen, obtengan, adquieran, transformen o controlen.

* Información pública

 Es toda información que un sujeto obligado genere, obtenga, adquiera, o controle en su calidad de tal.

* Información pública clasificada

Es aquella información que pertenece al ámbito propio, particular y privado o semiprivado de una persona natural o jurídica por lo que su acceso podrá ser negado o exceptuado, siempre que el acceso pudiere causar un daño a los siguientes derechos:

* + El derecho de toda persona a la intimidad, bajo las limitaciones propias que impone la condición de servidor público, en concordancia con lo estipulado.
	+ El derecho de toda persona a la vida, la salud o la seguridad.
	+ Los secretos comerciales, industriales y profesionales.

Estas excepciones tienen una duración ilimitada y no deberán aplicarse cuando la persona natural o jurídica ha consentido en la revelación de sus datos personales o privados o bien cuando es claro que la información fue entregada como parte de aquella información que debe estar bajo el régimen de publicidad aplicable.

* Información pública reservada

Es aquella información, cuyo acceso podrá ser rechazado o denegado de manera motivada y por escrito en las siguientes circunstancias, siempre que dicho acceso estuviere expresamente prohibido por una norma legal o constitucional:

* + La defensa y seguridad nacional.
	+ La seguridad pública.
	+ Las relaciones internacionales.
	+ La prevención, investigación y persecución de los delitos y las faltas disciplinarias, mientras que no se haga efectiva la medida de aseguramiento o se formule pliego de cargos, según el caso.
	+ El debido proceso y la igualdad de las partes en los procesos judiciales.
	+ La administración efectiva de la justicia.
	+ Los derechos de la infancia y la adolescencia.
	+ La estabilidad macroeconómica y financiera del país.
	+ La salud pública.

Se exceptúan también los documentos que contengan las opiniones o puntos de vista que formen parte del proceso deliberativo de los servidores públicos.

* + 1. **Licencias**

De acuerdo a la Guía de Datos Abiertos en Colombia, los datos abiertos están disponibles públicamente bajo una licencia abierta, la cual “es aquella que permite de forma libre y sin restricciones el uso, extracción, adaptación, transformación, copia, distribución, redistribución, reutilización, modificación, compilación y difusión de los datos publicados”.

Conforme a lo anterior, “se podrá hacer uso, transformación, distribución, redistribución, reutilización, compilación, extracción, copia, difusión, modificación y/o adaptación de los datos y de la información publicada en este sitio web, citando siempre la fuente de los datos. La entidad productora de los datos no será responsable de la utilización que hagan las personas que transformen y/o usen dichos datos, ni tampoco de los 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 licencia de uso se rige por la legislación colombiana, independientemente del entorno jurídico del usuario, cualquier disputa que llegue a surgir en la interpretación de estos términos se resolverá bajo el amparo de la Ley colombiana”.

Creative Commons (Commons, s.f.) cuenta con cuatro licencias aplicables a los datos abiertos:

* + Creative Commons Zero (CC0): ningún tipo de restricción, renunciando el creador a cualquier derecho sobre la obra, según permita la ley.
	+ Creative Commons Reconocimiento (CC BY): solo requiere que el reutilizador haga referencia a la fuente o autor original.
	+ Creative Commons Reconomiento-NoDerivadas (CC BY-ND): permite la reutilización y difusión siempre y cuando se haga referencia al autor de la misma y la obra no sufra cambio o alteración alguna.
	+ Creative Commons Reconocimiento-NoComercial (CC BY-NC): es posible alterar o difundir la obra original siempre y cuando se haga referencia al autor de la misma, careciendo de fines comerciales. La obra derivada no está obligada a mantener la misma licencia que la obra original.

En 2009, la organización OKFN (Open Knowledge Foundation Network) desarrolló las primeras licencias específicas para datos abiertos, las licencias Open Data Commons: (OKNF, s.f.)

* + Open Data Commons Public Domain Dedication and License (PDDL): que permite difundir, reutilizar o adaptar los datos sin restricción alguna.
	+ Open Data Commons Attribution License: al igual que la licencia de Creative Commons Reconocimiento, se exige la referencia a la autoría o fuente de los datos para la reutilización de la información.
	+ Open Data Commons Open Database License (ODbL): que permite la reutilización de los datos siempre que se reconozca la autoría de la información original; se mantenga la misma licencia en las obras derivadas las cuales puede restringir su uso si, además, se distribuye una versión sin dichas restricciones de uso.

**REFERENCIAS**

Berners-Lee, T. (febrero de 2009). *The next web*. Recuperado el mayo de 2017, de TED Ideas worth spreading: https://www.ted.com/talks/tim\_berners\_lee\_on\_the\_next\_web

Colombia. (1991). *CONSTITUCIÓN POLÍTICA DE COLOMBIA.* Recuperado el 20 de enero de 2018, de http://www.alcaldiabogota.gov.co/sisjur/normas/Norma1.jsp?i=4125

Colombia, D. A. (2017). *Portal Datos Abiertos Colombia*. Recuperado el mayo de 2017, de Ministerio de Tecnologías de la Información y las Comunicaciones MINTIC: https://www.datos.gov.co/

Commons, C. (s.f.). *Creative Commons Colombia.* Recuperado el 21 de enero de 2018, de Licencias: https://co.creativecommons.org/?page\_id=13

Congreso de Colombia. (16 de junio de 2001). *Ley 450 de 2011.* Recuperado el 20 de enero de 2018, de http://www.alcaldiabogota.gov.co/sisjur/normas/Norma1.jsp?i=43101

Congreso de Colombia. (30 de julio de 2009). *Ley 1341 de 2009.* Recuperado el 20 de enero de 2018, de http://www.mintic.gov.co/portal/604/articles-3707\_documento.pdf

Congreso de Colombia. (17 de octubre de 2012). *Ley 1581 de 2012.* Recuperado el 20 de enero de 2018, de http://www.alcaldiabogota.gov.co/sisjur/normas/Norma1.jsp?i=49981

Congreso de Colombia. (30 de junio de 2015). *Ley 1755 de 2015.* Recuperado el 20 de enero de 2018, de http://www.alcaldiabogota.gov.co/sisjur/normas/Norma1.jsp?i=62152

Congreso de la República. (31 de diciembre de 2008). *Ley 1266 de 2008.* Recuperado el 20 de enero de 2018, de http://wp.presidencia.gov.co/sitios/normativa/leyes/Documents/Juridica/Ley%201266%20de%2031%20de%20diciembre%202008.pdf

Congreso de la República. (18 de enero de 2011). *Ley 1437 de 2011.* Recuperado el 20 de enero de 2018, de http://www.secretariasenado.gov.co/senado/basedoc/ley\_1437\_2011.html

Congreso de la República. (12 de julio de 2012). *Ley 1564 de 2012.* Recuperado el 20 de enero de 2018, de http://www.alcaldiabogota.gov.co/sisjur/normas/Norma1.jsp?i=48425

Congreso de la República. (6 de marzo de 2014). *Ley 1712 de 2014.* Recuperado el mayo de 2017, de http://wsp.presidencia.gov.co/Normativa/Leyes/Documents/LEY%201712%20DEL%2006%20DE%20MARZO%20DE%202014.pdf

Congreso de la República de Colombia. (09 de junio de 2015). *Ley 1753 de 2015.* Recuperado el 20 de enero de 2018, de http://www.alcaldiabogota.gov.co/sisjur/normas/Norma1.jsp?i=61933

CONPES. (9 de febrero de 2000). *Agenda de Conectividad* . Recuperado el mayo de 2017, de Departamento Nacional de Planeación República de Colombia: https://www.mintic.gov.co/portal/604/articles-3498\_documento.pdf

El Pilón. (11 de noviembre de 2016). *En Tamalameque trabajan para mejorar calidad del agua.* Recuperado el mayo de 2017, de http://elpilon.com.co/tamalameque-trabajan-mejorar-calidad-del-agua/

El Tiempo. (5 de abril de 2016). *http://www.eltiempo.com/archivo/documento/CMS-16555655.* Recuperado el mayo de 2017, de http://www.eltiempo.com/archivo/documento/CMS-16555655

Few. (8 de noviembre de 2012). *Telling compelling stories with numbers.* Recuperado el mayo de 2017, de Perceptual Edge: http://www.actuate.com/download/acd2012/Telling-Compelling-Stories-with-Numbers.pdf

Few, S. (4 de mayo de 2017). *What Is Data Visualization?* Recuperado el mayo de 2017, de Visual Business Intelligence: http://www.perceptualedge.com/blog/?p=2636

Handbook, O. D. (s.f.). *What is Open Data?* Recuperado el mayo de 2017, de Open Knowledge Internacional: http://opendatahandbook.org/guide/en/what-is-open-data/

MINTIC. (12 de febrero de 2014). *Acerca del MinTIC / Historia*. Recuperado el mayo de 2017, de http://www.mintic.gov.co/portal/604/w3-propertyvalue-6077.html

MINTIC. (3 de agosto de 2015). *Datos Abiertos*. Recuperado el mayo de 2017, de Portal de Datos Abiertos, promoción de los datos y definición del Ecosistema de Datos Abiertos para Colombia: http://www.mintic.gov.co/portal/604/w3-article-7244.html

MINTIC. (2016). *Datos Abiertos. Participación, Colaboración y Transparencia.* Recuperado el mayo de 2017, de http://estrategia.gobiernoenlinea.gov.co/623/articles-9337\_presentacion\_datos\_abiertos.pdf

MINTIC. (01 de junio de 2016). *Guía de Datos Abiertos en Colombia.* Recuperado el mayo de 2017, de Gobierno en línea: http://estrategia.gobiernoenlinea.gov.co/623/articles-9407\_Guia\_Apertura.pdf

MINTIC Y FINDETER. (septiembre de 2016). *¿Qué es Emprende con datos?* Recuperado el mayo de 2017, de Emprende con Datos: http://emprendecondatos.gov.co/que-es/

MinTIC, M. d. (31 de diciembre de 2015). *Resolución 3564 de 2015.* Recuperado el 20 de enero de 2018, de http://www.mintic.gov.co/portal/604/articles-14476\_documento.pdf

Obama, B. (21 de enero de 2009). *Memorandum on Transparency and Open Government.* Recuperado el mayo de 2017, de Archivo nacional de los Estados Unidos: https://www.archives.gov/files/cui/documents/2009-WH-memo-on-transparency-and-open-government.pdf

OKFN. (16 de diciembre de 2015). *Global Open Data Index 2015*. Recuperado el mayo de 2017, de http://2015.index.okfn.org/place/

OKFN. (10 de noviembre de 2016). *Global Open Data Index 2016*. Recuperado el mayo de 2017, de https://index.okfn.org/place/

OKNF, O. K. (s.f.). *Open Data Commons.* Recuperado el 21 de enero de 2018, de Licenses: https://opendatacommons.org/licenses/

Open Data Charter. (s.f.). *Principios*. Recuperado el mayo de 2017, de Carta Internacional de Datos Abiertos: http://opendatacharter.net/principles-es/

Presidencia, R. d. (14 de abril de 2008). *DECRETO 1151 DE 2008.* Recuperado el mayo de 2017, de Lineamientos generales de la Estrategia de Gobierno en Línea: http://www.alcaldiabogota.gov.co/sisjur/normas/Norma1.jsp?i=29774

Presidente de la República de Colombia. (12 de diciembre de 2014). *Decreto 2753 de 2014.* Recuperado el 20 de enero de 2018, de http://www.mintic.gov.co/portal/604/articles-14673\_documento.pdf

Presidente de la República de Colombia. (26 de mayo de 2015). *Decreto 1074 de 2015.* Recuperado el 20 de enero de 2018, de http://wp.presidencia.gov.co/sitios/normativa/decretos/2015/Decretos2015/DECRETO%201074%20DEL%2026%20DE%20MAYO%20DE%202015.pdf

Presidente de la República de Colombia. (26 de mayo de 2015). *Decreto 1081 de 2015.* Recuperado el 20 de enero de 2018, de http://es.presidencia.gov.co/normativa/normativa/Decreto-1081-2015.pdf

Quintero, D. V. (13 de noviembre de 2016). *¿qué son los datos abiertos?* Recuperado el mayo de 2017, de Vive Digital TV: https://www.youtube.com/watch?v=-FpTIJFcVnE

W3C. (8 de septiembre de 2009). *Publishing Open Government Data*. Recuperado el mayo de 2017, de https://www.w3.org/TR/gov-data/

W3C. (s.f.). *Tim Berners-Lee Biography*. Recuperado el mayo de 2017, de https://www.w3.org/People/Berners-Lee/

1. La frase original es: “Big data, but little information”. [↑](#footnote-ref-1)