

2016

Datos Abiertos de Panamá

datosabiertos.gob.pa

“Más allá de la simplificación de los trámites – gobierno electrónico, y la transparencia – apertura parcial en cumplimiento de la ley, los datos abiertos reflejan una efectiva visibilidad y rendición de cuentas institucional; abren todo un universo a los ciudadanos, capaces de aprovechar esta oportunidad para crear nuevos productos de información y aplicaciones en favor de la sociedad”.



Guía de Implementación Técnica de Datos Abiertos de Gobierno

**“Una Estrategia de Apertura, Publicación y
Difusión”**

De datos públicos a datos abiertos.

*Los organismos estatales recogen, producen, reproducen y difunden datos para cumplir con su función pública. Algunos de estos datos ya están publicados o debieran estarlo. Incorporar la publicación de **datos públicos en formatos abiertos**, abre la puerta a la posibilidad que los mismos sean reutilizados en nuevos proyectos, que puedan combinarse con otras fuentes de datos y que puedan generar nuevas aplicaciones desarrolladas por el gobierno, por la sociedad civil, organizaciones, empresas o ciudadanos en general.*

Esta guía brinda introducción al tema de datos abiertos. La misma reúne conceptos básicos de datos abiertos de gobierno, los pasos a seguir a la hora de abrir los datos, ejemplos de reutilización de datos y de herramientas para su procesamiento. También se ofrecen referencias a contenidos más específicos para el lector que quiera profundiza en el tema.

BUENAS PRÁCTICAS Versión 1.0 – 2016

Este documento ha sido adaptado (2016) por la Autoridad Nacional para la Innovación Gubernamental (AIG) de la República de Panamá, con el respaldo de la Agencia para el Desarrollo del Gobierno de Gestión Electrónica y la Sociedad de la Información y el Conocimiento (AGESIC), de la República Oriental de Uruguay.

Usted es libre de copiar, distribuir, comunicar y difundir públicamente este documento así como hacer obras derivadas, siempre y cuando tengan en cuenta citar la obra de forma específica y no utilizar esta obra para fines comerciales. Toda obra derivada de esta deberá ser generada con estas mismas condiciones.

CONTENIDO

INTRODUCCIÓN.....	5
POTENCIAL Y BENEFICIOS DE LOS DATOS ABIERTOS DE GOBIERNO.....	5
CARACTERÍSTICAS QUE DEBEN TENER LOS DATOS ABIERTOS ¹.....	6
ESTRATEGIA PARA LA APERTURA Y PUBLICACIÓN DE DATOS ABIERTOS	7
I. PLAN INSTITUCIONAL PARA LA APERTURA DE DATOS	7
A. Enlace de Gobierno Abierto	7
B. Equipo de trabajo	7
C. Generación del Inventario Institucional de Datos	8
D. Priorizar los datos de valor	9
E. Crear una agenda institucional de apertura y publicación de datos abiertos y presentarla a la comunidad	9
II. PREPARACIÓN Y PUBLICACIÓN DE LOS DATOS ABIERTOS	10
¿Qué datos de gobierno podrían ser abiertos?	10
A. Preparación de los conjuntos de datos.....	11
1. Clasificación de los datos abiertos.....	11
EXPLICACIÓN DE LA CLASIFICACIÓN POR ESTRELLAS	11
1 Estrella.....	11
2 Estrellas	12
3 Estrellas	12
4 Estrellas	12
5 Estrellas	12
Enlaces	13
2. Preparación de los datos	14
a. Seleccionar el conjunto de datos a abrir.....	14
b. Convertir el formato del archivo a un formato abierto	14
-Desde un archivo de Excel.....	14
-Desde un gestor de bases de datos.....	16
c. Aplicar la licencia abierta	19
d. Incluir los metadatos de los archivos (documentación)	20
e. Empaquetar	21

f. Asignar URL o URI (caso de conjuntos de datos de 4 y 5 estrellas)	21
3. Publicación de los conjuntos de datos en el sitio Web institucional	22
4. Publicación de los conjuntos de datos en el Portal Nacional datoabiertos.gob.pa.....	22
5. Difusión y promoción de los datos abiertos	25
GLOSARIO DE TÉRMINOS	26

INTRODUCCIÓN

La modificación introducida al presente documento se refiere a la estrategia organizativa institucional y a la creación de los procesos institucionales previos a la publicación de los datos abiertos. Lo anterior se desprende de la necesidad de establecer las directrices generales para formalizar las operaciones relacionadas a la apertura y publicación de datos abiertos, en los sitios institucionales y portal nacional de datos abiertos.

Las instituciones públicas del estado recopilan y producen grandes volúmenes de datos relacionados con energía, clima, economía, salud, medio ambiente, agricultura, defensa, seguridad pública, sociedad, cultura, presupuestos, entre muchos otros.

El hecho de que los ciudadanos puedan acceder a todos los datos de las instituciones públicas en forma simple impactará positivamente en nuestra sociedad, mejorará la democracia, producto de mayores niveles de participación y compromiso sociedad-gobierno, y va a mejorar la economía ya que se ha demostrado que el acceso a datos públicos permite realizar obras derivadas a partir de esos datos con su consecuente impacto en la innovación y el emprendimiento.

“Los datos abiertos son aquellos que se emiten en formato primario (sin procesar), que pueden ser utilizados, reutilizados y redistribuidos libremente por cualquier ciudadano sin restricciones, y que se encuentran sujetos, cuando más, al requerimiento de reconocimiento y de compartirlos de la misma manera en que aparecen”.

POTENCIAL Y BENEFICIOS DE LOS DATOS ABIERTOS DE GOBIERNO

Los datos generados por las instituciones públicas al ser puestos a disposición ciudadana como Datos Abiertos, incrementan su potencial de uso, de reutilización y de redistribución, en el sentido de:

- ❖ Impulsar el crecimiento económico, fortalecer la competitividad y promover la innovación;
- ❖ Incrementar la transparencia y rendición de cuentas;
- ❖ Fomentar la participación ciudadana y fortalecer la democracia;
- ❖ Estimular una mayor eficiencia gubernamental, lo que obliga a diseñar políticas públicas en base a las necesidades de la comunidad y a mejorar los servicios públicos, en apoyo a los objetivos de desarrollo, generación de conocimiento y buena gobernanza;
- ❖ Emitir datos en el marco del respeto a la privacidad y protección de la información personal.
- ❖ Permitir la interoperabilidad entre organizaciones y sistemas. La habilidad de integrar componentes es esencial para construir sistemas complejos, grandes y ricos en información.
- ❖ Fomentar la investigación y el periodismo de datos, lo que permite recabar y analizar grandes cantidades de datos mediante software especializado para hacer comprensible la información a la comunidad a través de artículos, infografías, y visualizaciones de datos, entre otros.

CARACTERÍSTICAS QUE DEBEN TENER LOS DATOS ABIERTOS ¹

1. **Accesibles:** Disponibles en línea y de fácil acceso;
2. **Integrales:** Deberán ser completos y sin procesar, en la medida de lo posible deben contener el tema que describen a detalle y con los metadatos necesarios;
3. **Gratuitos:** Se obtendrán sin entregar a cambio contraprestación alguna; tomando en cuenta lo establecido en la ley 6 de enero de 2002;
4. **No discriminatorios:** Serán accesibles sin restricciones de acceso para los usuarios;
5. **De libre uso:** las entidades indicarán la licencia más apropiada que aplique a los datos liberados, como ejemplo la licencia CC BY 4.0, que requiere citar la fuente de origen como único requisito;
6. **Legibles por máquinas:** Deberán estar estructurados, total o parcialmente, para ser procesados e interpretados por equipos electrónicos de manera automática;
7. **Primarios:** Provendrán de la fuente de origen con el máximo nivel de granularidad posible, no modificados y sin procesar;
8. **Oportunos:** Serán actualizados periódicamente, conforme se generen o cambien;
9. **Permanentes:** Se deberán conservar en el tiempo, para lo cual, las versiones históricas relevantes para uso público, se mantendrán disponibles a través de identificadores adecuados para tal efecto.

¹ <http://www.opengovdata.org/home/8principles>

ESTRATEGIA PARA LA APERTURA Y PUBLICACIÓN DE DATOS ABIERTOS

I. PLAN INSTITUCIONAL PARA LA APERTURA DE DATOS

A. Enlace de Gobierno Abierto

Establecer el responsable institucional de los datos abiertos. Este responsable será el *enlace de gobierno abiert*, entre cuyas funciones estarán:

1. Convocar a los futuros actores de la apertura de datos y conformar el equipo de trabajo;
2. Coordinar con los integrantes del equipo de trabajo y vigilar la gestión del mismo en la correcta implementación de las directrices para la apertura y publicación de datos;
3. Conformar el plan, junto con el equipo de trabajo, de la apertura de datos, definiendo las actividades y tareas, igualmente, delegando funciones a los miembros del equipo;
4. Coordinar la creación del inventario institucional de datos;
5. Priorizar, en conjunto con el equipo y de acuerdo a consultas a la sociedad civil, los datos con la mayor demanda ciudadana;
6. Velar por la aplicación de la norma legal vigente en la publicación de datos personales, confidenciales o con derecho de autor (operación de filtrado);
7. Implementar los procesos para que los datos se publiquen en el sitio institucional y en el portal nacional;
8. Ser el enlace institucional con la AIG para la coordinación y actualización del portal nacional de datos abiertos ***datosabiertos.gob.pa***;
9. Llevar el control del cumplimiento del plan institucional para la apertura de datos, de acuerdo al avance de las actividades y tareas realizadas por el grupo de trabajo.

B. Equipo de trabajo

Crear el equipo de trabajo con las personas involucradas tanto en la generación de los datos abiertos y su mantenimiento permanente, como en la actualización del sitio Web institucional y el portal nacional de datos abiertos. El equipo de trabajo actuará como un consejo institucional de datos abiertos para coordinar la implementación de la política de datos abiertos en toda la institución; el equipo deberá estar formado como mínimo por:

1. El administrador de datos abiertos y enlace institucional;
2. La Dirección de Informática;
3. Los departamentos, unidades y áreas que generan y/o administran datos;

4. El enlace con la Autoridad Nacional de Transparencia y Acceso a la Información (ANTAI);
5. El asesor legal, como apoyo para analizar, asesorar y orientar sobre los datos publicados, y que se respete la reserva legal de los documentos que la tienen.

Entre las funciones del equipo de trabajo estarán las siguientes:

1. Acordar compromisos para el desarrollo de las fases de apertura de datos con todas las áreas involucradas de la entidad;
2. Planificar y documentar las actividades, tareas y estimaciones de tiempo necesarias para la apertura de datos, entre las que identificamos las principales:
 - Conformación del inventario institucional de datos
 - Priorización de los datos a publicar
 - Selección de los conjuntos de datos a liberar
 - Conversión a los formatos abiertos
 - Presentación en el sitio Web de la institución y portal nacional ***datosabiertos.gob.pa***
3. Organizar todos los recursos que sean necesarios para la apertura de los primeros conjuntos de datos;
4. Ejecutar las actividades y tareas necesarias dirigidas a la presentación en el sitio Web institucional y portal nacional, de los datos en formato abierto.

C. **Generación del Inventario Institucional de Datos**

Para crear el inventario institucional de datos es necesario incluir toda la información que produce la institución, a partir de las bases de datos y otros sistemas de manejo de datos donde se incluya información estadística, financiera, investigativa, administrativa, legal, de geolocalización, de proyectos, entre otros. El administrador de datos solicitará a los responsables de generar y/o administrar datos, el llenado de la plantilla del inventario institucional de datos para cada conjunto de datos o servicio de información que generan y/o administran.

El inventario de datos será conformado de manera inicial y continua, con el insumo que proporcionen las unidades administrativas y áreas responsables de generar y/o administrar datos, las cuales deberán reportar los datos que generan al administrador de datos, responsable de consolidar la información en un solo listado que posteriormente se publicará en el catálogo de datos abiertos en ***datosabiertos.gob.pa***. El listado se puede constituir con los campos mostrados en la tabla abajo:

Responsable del conjunto de datos o servicio de información	Nombre del conjunto de datos o servicio de información	Temas que lo componen	Descripción del servicio o conjunto de datos	Contiene datos privados, confidenciales o de seguridad - describir	Plataforma, tecnología, aplicación o programa en que están los datos	Período de actualización del conjunto de datos o servicio de información	Estado del conjunto de datos con respecto a sus características

D. Priorizar los datos de valor

Es importante decidir el orden de publicación de los datos abiertos, y la prioridad la va a determinar el potencial de los mismos en generar impacto en la comunidad. Los datos que generan gran impacto son aquellos que tienen mayor valor para la comunidad. El grupo de trabajo deberá priorizar los conjuntos de datos del inventario institucional de datos, según los siguientes criterios:

- Considerando los conjuntos de datos de interés y demandados por la comunidad, de acuerdo a la experiencia con las solicitudes de información recurrentes realizadas por esta;
- Considerando todos aquellos conjuntos de datos de carácter público que normalmente son publicados en el sitio Web y otros esquemas institucionales;
- Solicitando a la persona que es el enlace con Transparencia un reporte de las solicitudes de acceso a la información pública realizadas con mayor frecuencia;
- Teniendo en cuenta el estado de los conjuntos de datos, es decir, deben cumplir con las características descritas arriba para considerarlos datos abiertos.

E. Crear una agenda institucional de apertura y publicación de datos abiertos y presentarla a la comunidad

Los resultados de los puntos anteriores A-D, que conforman el Plan Institucional de Apertura de Datos, deberán ser publicados en su sitio institucional de datos abiertos, de acuerdo al siguiente formato:

Administrador de Datos Abiertos	Título del Administrador de Datos Abiertos	Correo oficial del Administrador de DA	Conjuntos de datos priorizados para la liberación	Descripción de los conjuntos de datos	Periodicidad de la actualización de DA	Fecha tentativa de publicación (DD-MM-AAAA)
---------------------------------	--	--	---	---------------------------------------	--	---

Este plan será publicado anualmente con el objetivo de dar a conocer las acciones que serán realizadas por la institución en materia de datos abiertos. Como uno de sus componentes principales, el plan de apertura contendrá el calendario de publicación anual de los datos abiertos de la Institución.

II. PREPARACIÓN Y PUBLICACIÓN DE LOS DATOS ABIERTOS

Para la preparación y publicación de datos es necesario considerar los siguientes puntos sobre los cuales se debe trabajar:

- ❖ Preparación de los conjuntos de datos;
- ❖ Publicación de los datos abiertos en el sitio Web institucional;
- ❖ Documentación de los conjuntos de datos publicados;
- ❖ Publicación del catálogo institucional de datos abiertos en el portal nacional *datosabiertos.gob.pa*

¿Qué datos de gobierno podrían ser abiertos?

“Los datos públicos de gobierno, según se definen en la ley de acceso a la información pública”. El punto clave a destacar es que la apertura de datos se centra en los datos no personales, los que no están sujetos a restricciones específicas. Como ejemplos tenemos el presupuesto público, ejecución del presupuesto, estadísticas, datos del censo, compras públicas, desempeño de servicios, datos del transporte, datos judiciales/crímenes, contaminación y energía, comercio internacional, datos del clima, y otros.

Empezando con lo más sencillo, comparta los datos que ya tiene, seleccionando los que son fáciles de publicar y asegurándose que los mismos tienen alta demanda dentro de la comunidad. Prepare series de conjuntos de datos, por ejemplo series anuales, de tal forma que se puedan bajar datos completos masivamente para los análisis. Publique en formatos abiertos.

Al utilizar formatos abiertos, se amplía la gama de usuarios que pueden aprovechar los datos, ya que estos formatos permiten que sean interpretados por la mayoría de los sistemas computacionales. Es altamente recomendable que los datos se publiquen en formatos estructurados como el CSV. Sin embargo, para dar respuesta a toda esa gama de usuarios, sectores y casos de uso, se debe considerar la publicación de los mismos conjuntos en diferentes formatos estructurados de estándar abierto convenientes al tipo de usuario, como lo son XML, JSON, RDF, GeoJSON, KML, DBF, incluyendo formatos como SHP y XLSX. En general, se recomienda la publicación de datos en formatos altamente estructurados y abiertos;

Los conjuntos de datos deben estar disponibles para descarga directa en un URL único y sin registro de acceso mediante contraseñas o interfaces de usuario, ya que estos datos serán procesados por computadoras, no sólo por personas. Para lo anterior, se recomienda cargar los conjuntos de datos abiertos en el servidor web de la institución en la sección destinada a mostrar la información institucional, para que paulatinamente se vaya reemplazando los datos en formato tradicional, por datos en formatos abiertos, además de incluir conjuntos de datos nuevos. Para facilitar las descargas de los datos abiertos, los archivos muy grandes se deben segmentar y comprimir (utilizando formatos abiertos de compresión como gzip). Los URLs incluidos en los conjuntos de datos publicados en el

sitio Web de la institución, serán utilizados posteriormente para referenciar estos conjuntos de datos en el Catálogo de Datos Abiertos en ***datosabiertos.gob.pa***.

A. Preparación de los conjuntos de datos

Para mayor claridad sobre el formato de los datos a publicar, la siguiente clasificación permite orientarse sobre los niveles de apertura que podemos dar a los datos.

1. Clasificación de los datos abiertos

Tim Berners-Lee² propuso una categorización del grado de apertura de los datos basado en estrellas, refiriéndose a qué tan abiertos y usables son los datos que puede ofrecer una institución:

1 estrella: Los datos deben estar disponibles en la web en cualquier formato y bajo una licencia abierta, para que sean considerados como datos abiertos.

2 estrellas: Los datos deben estar estructurados y en un formato que pueda ser interpretado o procesado por máquinas; deberán estar disponibles en la Web y tener una licencia abierta de uso.

3 estrellas: Igual al anterior pero en un formato no propietario.

4 estrellas: Todo lo anterior más la utilización de estándares recomendados por w3c (RDF – SPARQL) para identificar cosas.

5 estrellas: Todo lo anterior más el establecimiento de vínculos con los datos de otros y de esa forma proveerles contexto.

EXPLICACIÓN DE LA CLASIFICACIÓN POR ESTRELLAS

1 Estrella

Según la clasificación antes descrita, cualquier dato que esté disponible en la Web en cualquier formato bajo una licencia abierta, es un dato abierto de categoría 1 estrella.

Más adelante se explica lo que implica una licencia abierta, pero a grandes rasgos es un documento que expresa formal y explícitamente las condiciones bajo las cuales podrán ser utilizados los datos presentados en un sitio Web.

Cualquier archivo disponible en sitios Web en formato .PDF, .DOC, .ODT, .TXT, entre otros, y que esté bajo una licencia abierta se considera dato abierto de 1 estrella. En

² <http://www.w3.org/People/Berners-Lee/>

cualquiera de los formatos anteriores el archivo se debe poder imprimir, guardar en el disco duro, copiar, usar los datos en otros sistemas, compartirlos. Los datos en este nivel son fáciles de publicar.

2 Estrellas

El archivo tiene que cumplir todo lo que implica estar en 1 estrella (publicado y bajo una licencia abierta) y además los datos deben estar estructurados y en un formato que pueda ser interpretado o procesado por maquinas.

Por ejemplo, los datos que se encuentren en formato .XLS, y estén organizados en una tabla, sin contraseña para su uso se pueden clasificar como 2 estrellas. Los archivos en este nivel pueden ser procesados con software propietario, para agregarlos, hacer cálculos o exportarlos a otro formato estructurado. También es fácil publicar este tipo de archivos.

3 Estrellas

El nivel 3 estrellas implica lo mismo que los niveles 1 y 2 pero el formato debe ser no propietario.

Por ejemplo, datos que se encuentren en formato .CSV son de nivel 3 estrellas. Otro ejemplo es un archivo en formato .XML. Se puede manipular los datos de cualquier forma sin limitación de características o de uso de algún tipo de software en particular para ello. Son fáciles de publicar.

4 Estrellas

Para que un dato abierto se considere de 4 estrellas debe cumplir con las condiciones anteriores y además utilizar estándares recomendados por W3C como RDF y SPARQL. En este nivel los datos se pueden vincular desde cualquier otro sitio (Web o local); se pueden marcar como favoritos; reutilizar partes de los datos como también las herramientas y librerías disponibles, incluso si estas solo entienden parte de los patrones que utilizó quien los publicó. Entender la estructura de un gráfico de datos RDF puede requerir más esfuerzo que el requerido para entender datos tabulares (Excel/CSV) o en estructura de árbol (XML/JSON). Se debe usar los URIs para denotar cosas, así los usuarios pueden apuntar a estas.

5 Estrellas

En este nivel se enlazan los datos cargados a otros datos para proveer contexto. Las ventajas en este nivel es que se puede descubrir más datos (relacionados) mientras se está utilizando la información, como también, se puede aprender directamente acerca del esquema de datos. El publicador puede hacer que los datos sean descubiertos; se puede incrementar el valor de los datos; se ganan los mismos beneficios de los enlaces como los consumidores. Sin embargo, se necesitará invertir recursos para enlazar los datos a otros en la Web y podría haber necesidad de tener que reparar enlaces rotos o incorrectos.

Los niveles 4 y 5 son considerados avanzados y se acercan mucho al ideal de publicación de un dato abierto. En la mayoría de los casos, se espera iniciar la liberación de datos de gobierno utilizando como mínimo el nivel de 3 estrellas. Una vez afianzado el proceso de

apertura, tanto organizativa como operativamente, se espera que la institución incluya los niveles 4 y 5 estrellas en su gestión de liberación de datos, para posibilitar un uso más amplio de la información.

Enlaces

Para aquellos lectores que estén interesados en profundizar en el tema, a continuación se listan enlaces con más información:

- **RDF, SPARQL y Linked Data**

Guías y tutoriales.

<http://linkeddata.org/guides-and-tutorials>

<http://www.w3.org/TR/rdf-sparql-query/>

- **Comunidad de Web Semántica**

Uno de los principales lugares para obtener acceso a información y recursos sobre web semántica es el sitio de esta comunidad.

http://semanticweb.org/wiki/Main_Page

- **Para desarrolladores PHP**

Graphite es una librería para PHP construida sobre RC2 y facilita el desarrollo de aplicaciones que utilicen datos en formato RDF. Su nueva funcionalidad de “descripción de recursos” la hace una excelente forma de proveer a un sitio web datos desde SPARQL.

<http://graphite.ecs.soton.ac.uk/>

- **Para desarrolladores Java**

<http://jena.apache.org/>

<http://www.openrdf.org/>

- **Para desarrolladores .NET**

<http://www.brightstardb.com/>

<http://code.google.com/p/lingtordf/>

- **Implementaciones SPARQL**

<http://www.w3.org/wiki/SparqlImplementations>

- **Protégé**

Es un editor de ontologías gratuito y de código abierto.

<http://protege.stanford.edu/>

2. Preparación de los datos

Como guía para la preparación de los datos abiertos de su institución recomendamos seguir los siguientes pasos:

- a) Seleccionar el conjunto de datos a abrir
- b) Convertir el formato del archivo a un formato abierto
- c) Aplicar la licencia abierta
- d) Incluir los metadatos
- e) Empaquetar
- f) Para los conjuntos de datos de 4 y 5 estrellas asignar un URI (Uniform Resource Identifier o identificador uniforme de recursos, que sirve para identificar recursos en Internet, precisamente lo que el nombre indica).

A continuación se describe cada uno de estos pasos:

a. Seleccionar el conjunto de datos a abrir

Sin importar la clasificación, en todos los casos, el primer paso es seleccionar el conjunto de datos para liberar. Por ello hay que comenzar a identificar los datos que ya tenemos publicados o que son de fácil transformación. Tome en cuenta que los conjuntos de datos deben ser completos, primarios, actualizados y anonimizados. Priorice respondiendo las siguientes preguntas:

- ¿Qué información le solicitan habitualmente y qué exige un procesamiento de datos?
- ¿Qué información existe en su oficina con datos que pueden ser de interés para la población y grupos de interés?
- ¿Qué información se brinda habitualmente a la prensa, o se publica en medios gráficos que puede ser reutilizada?
- ¿Qué información se brinda a otros organismos que puede ser de interés general?

Una vez que haya identificado la información candidata para publicar, debe seleccionar la que se encuentre en condiciones de ser publicada. Tenga en cuenta aspectos legales, de completitud, capacidad para mantenerla actualizada, formato, elimine los datos personales si existen (anonimización), etc. Si piensa publicar datos utilizando formatos de 1 estrella (.PDF, .DOC, .ODT, .TXT), cerciórese que no llevan contraseñas contra lectura, impresión, copia, descarga, etc.

b. Convertir el formato del archivo a un formato abierto

Si tiene planeado publicar datos con más de 1 estrella, en este punto mostraremos 2 casos en que se transforma el formato del conjunto de datos a publicar. Esto posibilita mejorar la clasificación del dato abierto. Por ejemplo, si se está en 1 o 2 estrellas podría pasar a 2 o 3 estrellas respectivamente.

La conversión se puede hacer desde un archivo o desde un gestor de base de datos; comenzaremos por la primera.

-Desde un archivo de Excel

Ejemplo de cambio de formato de .XLS a .CSV

A modo de ejemplo tomamos información relacionada con un censo de hogares en formato

Excel y realizaremos la transformación a formato abierto CSV.

	A	B	C	D	E	F	G	H	I	J	K
1	CUADRO 1										
2	Población por departamento de residencia habitual, según área y sexo. Total País										
3											
4											
5	Departamento	Total	Hombres	Mujeres	Ignorado	Área Urbana					
6						(1)	Total	Hombres	Mujeres	Ignorado	
7						(1)					(1)
8	Total	3.286.314	1.577.725	1.708.481	108	3.110.701	1.478.967	1.631.626	108		
9											
10	Montevideo	1.319.108	613.990	705.014	104	1.305.082	604.965	700.013	104		
11											
12	Resto País	1.967.206	963.735	1.003.467	4	1.805.619	874.002	931.613	4		
13											

Para facilitar el procesamiento de los datos de manera automatizada se podría llevar esta planilla a un formato de texto separado por comas, incluso removiendo alguna información previamente de manera de hacer más fácil su reutilización.

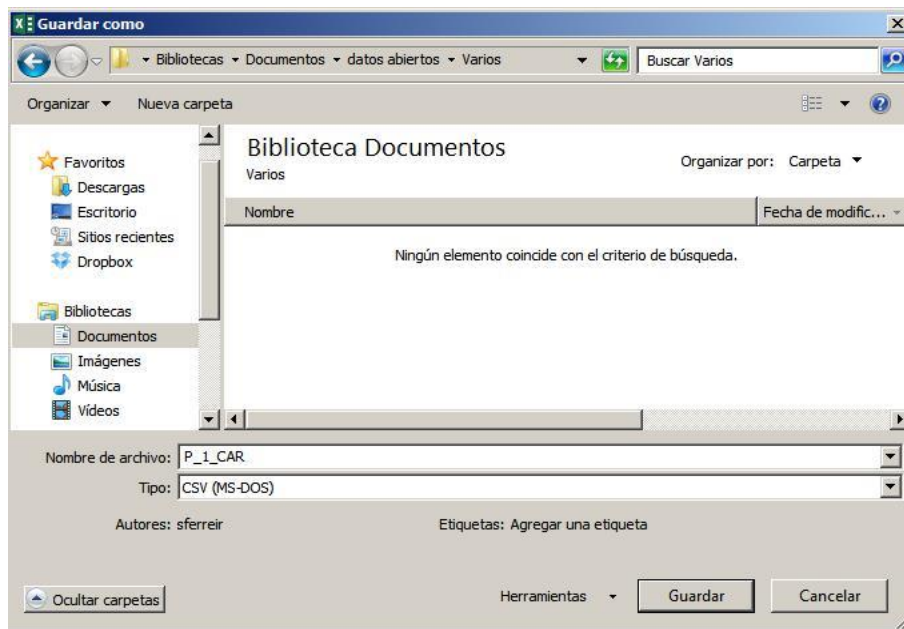
Como primer paso eliminamos la información innecesaria dejando solamente los nombres de las columnas, filas y los datos numéricos.

Departamento	Total	Hombres	Mujeres	Ignorado	Total Urbano	Hombres Urbano	Mujeres Urbano	Ignorado
Total	3.286.314	1.577.725	1.708.481	108	3.110.701	1.478.967	1.631.626	108
Montevideo	1.319.108	613.990	705.014	104	1.305.082	604.965	700.013	104
Resto País	1.967.206	963.735	1.003.467	4	1.805.619	874.002	931.613	4
Artigas	73.378	36.079	37.299	-	69.854	34.102	35.752	-

Luego guardamos el archivo como CSV (MS-DOS). Si después de guardado en formato CSV lo abrimos con un editor de texto, veremos lo siguiente:

```
Departamento;Total;;Hombres;Mujeres;Ignorado;;Total Urbano;Hombres Urbano;
Total;3.286.314;;1.577.725;1.708.481;108;;3.110.701;1.478.967;1.631.626;108
Montevideo;1.319.108;;613.990;705.014;104;;1.305.082;604.965;700.013;104;;1
Resto País;1.967.206;;963.735;1.003.467;4;;1.805.619;874.002;931.613;4;;161
Artigas;73.378;;36.079;37.299;-;;69.854;34.102;35.752;-;;3.524;1.977;1.547;
Canelones;520.187;;253.124;267.063;-;;471.968;227.538;244.430;-;;48.219;25.
Cerro Largo;84.698;;41.050;43.648;-;;78.762;37.724;41.038;-;;5.936;3.326;2.
```

Convertir un archivo de Excel (lo mismo ocurre desde Open Office o Libre Office) a formato CSV es muy sencillo, simplemente en la opción “Guardar como” del documento se debe seleccionar el tipo de archivo que queremos.



-Desde un gestor de bases de datos

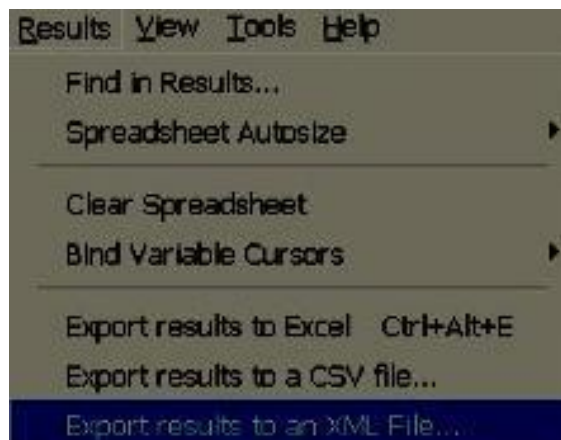
Para la extracción de datos desde bases de datos, se muestran ejemplos de cómo puede realizarse desde algunos de los gestores de bases de datos disponibles, al formato XML.

Ejemplo: exportación de datos desde Oracle al formato XML

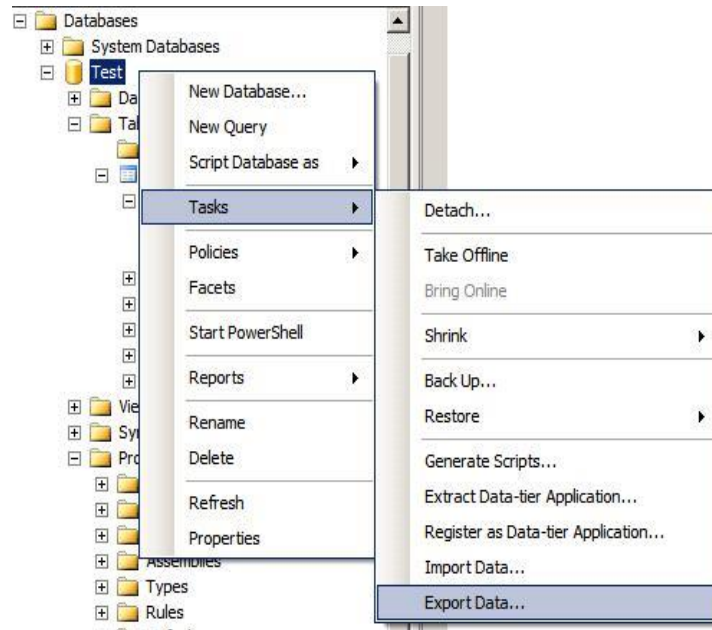
Oracle no cuenta con una interfaz gráfica que permita realizar una exportación de datos a XML directamente, es necesario crear un procedimiento que lea desde una estructura de tabla o vista y genere el documento XML para esos datos.

Sin embargo existen otras herramientas que permiten realizar la exportación, por ejemplo PL/SQL Developer permite realizar la exportación directa de datos a formato XML y CSV desde un menú como muestra la siguiente imagen:

Ejemplo: exportación de datos al formato XML mediante MSSQL Server



MSSQL Server cuenta con un asistente para la exportación de datos muy sencillo de utilizar, permite exportar a diferentes formatos como CSV o Excel. La siguiente es una imagen del menú mencionado:



En su versión 2008 R2 no es posible la exportación directa desde la consola de administración a un formato XML. Para exportar datos a formato XML es necesario realizar cierta programación de manera de lograr el formato deseado. Para ello el desarrollo de procedimientos es la manera indicada donde además se pueden utilizar funciones específicas para la generación de este tipo de documentos.

La versión 2012 de SQL Server provee un comando para la exportación de datos en formato XML que facilita la solución. El comando es el siguiente:

```
bcp tabla_o_vista format nul -f nombre_de_archivo.xml -x
```

Ejemplo: exportación de datos al formato XML mediante MySQL

Utilizando como ejemplo a MySQL podemos fácilmente exportar los datos de una tabla a formato XML válido. Desde la línea de comando se debe ejecutar lo siguiente:

```
>mysqldump --xml test test --user=root --password=mysql
```

Tabla denominada “test” exportada a XML:

Idcampo1	Campo2
1	registro 1
2	registro 2
3	registro 3
4	registro 4

La salida que se obtendrá será la estructura de la tabla, sus campos y sus datos. A continuación se transcribe la salida mencionada.

```
<?xml version="1.0"?>

<mysqldump          xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance">

  <database name="test">

    <table_structure name="test">

      <field Field="idcampo1" Type="int(11)" Null="NO" Key="PRI"
      Extra="" Comment="" />

      <field Field="campo2" Type="varchar(50)" Null="NO" Key="" Extra=""
      Comment="" />

      <key Table="test" Non_unique="0" Key_name="PRIMARY"
      Seq_in_index="1" Column_name="idcampo1" Collation="A"
      Cardinality="2" Null="" Index_type="BTREE" Comment=""
      Index_comment="" />

      <options Name="test" Engine="InnoDB" Version="10" Row_format="Compact"
      Rows="4" Avg_row_length="4096" Data_length="16384" Max_data_length="0"
      Index_length="0" Data_free="10485760" Create_time="2012-08-22
      21:24:11" Collation="latin1_swedish_ci" Create_options="" Comment=""
      />

    </table_structure>

    <table_data name="test">

      <row>

        <field name="idcampo1">1</field>

        <field name="campo2">registro 1</field>

      </row>

      <row>

        <field name="idcampo1">2</field>

        <field name="campo2">registro 2</field>

      </row>

      <row>

        <field name="idcampo1">3</field>

        <field name="campo2">registro 3</field>

      </row>

      <row>

        <field name="idcampo1">4</field>
```

```

<field name="campo2">registro 4</field>
</row>
</table_data>
</database>
</mysqldump>

```

c. Aplicar la licencia abierta

La licencia a aplicar a un conjunto de datos es el documento que expresa formal y explícitamente las condiciones bajo las cuales podrán ser utilizados los datos disponibles en un sitio Web.

El portal **datosabiertos.gob.pa** utilizará la licencia de la organización Creative Commons, de aplicación general conocida como “Reconocimiento 4.0 Internacional (CC BY 4.0)” para los conjuntos de datos contenidos en el Catálogo Nacional, cargados allí por las instituciones (ver http://creativecommons.org/licenses/by/4.0/deed.es_ES y su texto completo en <http://creativecommons.org/licenses/by/4.0/legalcode>). Es posible que alguna institución quiera, independientemente, aplicar otro tipo de licencia al conjunto de datos, como por ejemplo: “De Dominio Público” (ver https://creativecommons.org/publicdomain/zero/1.0/deed.es_ES y su texto completo en <https://creativecommons.org/publicdomain/zero/1.0/legalcode>), lo que es igualmente posible.

Parte del texto que contiene la licencia “Reconocimiento 4.0 Internacional (CC BY 4.0)”, es el siguiente:

Usted es libre de:

- ***Compartir*** — copiar y redistribuir el material en cualquier medio o formato
- ***Adaptar*** — remezclar, transformar y crear a partir del material para cualquier finalidad, incluso comercial.
- *El licenciador no puede revocar estas libertades mientras cumpla con los términos de la licencia.*

Reconocimiento — Debe ***reconocer adecuadamente*** la autoría, proporcionar un enlace a la licencia e ***indicar si se han realizado cambios***. Puede hacerlo de cualquier manera razonable, pero no de una manera que sugiera que tiene el apoyo del licenciador o lo recibe por el uso que hace.

¿Cómo se indica la licencia en el Catálogo Nacional de Datos Abiertos?

En el proceso de carga de datos al portal nacional se ofrece una entrada para la licencia. Puede hacer referencia a la licencia para datos abiertos que se encuentra publicada en el sitio <http://datosabiertos.gob.pa>. Por otro lado, se puede hacer uso de la licencia que acostumbra especificar en el sitio de su institución donde carga los datos abiertos, si es el caso.

d. Incluir los metadatos de los archivos (documentación)

Para facilitarles a los usuarios finales el uso de la información liberada es necesario hacer una descripción detallada de los datos que se publican, mediante los metadatos. Los metadatos hacen una descripción específica del conjunto de datos (Ejemplo: Nombre del conjunto de datos, descripción, palabras clave, información legal, licencia, cobertura temporal y geográfica, fecha de publicación, última actualización). Para el caso de estructuras relacionales u hojas de cálculo las columnas pueden ser los metadatos; para los archivos XML se puede considerar el DTD, etc.

Ejemplo: Incluir metadatos de un archivo en formato .XLS, .ODS o CSV

A la hora de suministrar los datos abiertos en un formato .XLS, .ODS o .CSV, también se debe publicar otro con la licencia (si aplica) y un tercer archivo con los metadatos (a descargar del sitio).

Si el archivo está en formato CSV, los metadatos deben explicar el significado de cada columna, como se muestra en la siguiente tabla:

Nombre de atributo	Descripción	Tipo de dato	Recurso relacionado	Información Adicional
<Descripción exacta de la columna tal cual aparece en los archivos de los conjuntos de datos>. Ejemplo: CodigoProv	<Descripción detallada del significado del atributo orientada a las personas que pretendan utilizar el atributo>. Ejemplo: Provincia de Residencia	<Numérico, Texto, Fecha/Hora, Moneda> Ejemplo: Numérico	<Recurso que da significado al atributo. Ejemplo: si el atributo es un código tiene como recurso relacionado la tabla o recurso donde se describe el significado de dicho código.> Ejemplo:Provincia.xls	<Cualquier información adicional sobre el atributo que complemente la información básica.> Ejemplo: En el archivo <i>Provincia.xls</i> están los nombres de las Provincias asociadas a dicho código. Ejemplo: 1- Panamá; 2-Colón; 3-Coclé....

e. Empaquetar

Una vez que se dispone del archivo de datos en un formato abierto, un archivo con la licencia (en el caso que aplicara y no se usara la provista por defecto del sitio donde se está publicando) y un archivo con los METADATOS del conjunto de datos (usualmente: Nombre, descripción, palabras clave, información legal, licencia, cobertura temporal y geográfica, fecha de publicación, última actualización), se debe empaquetar toda la información utilizando herramientas para la compresión de archivos.

f. Asignar URL o URI (caso de conjuntos de datos de 4 y 5 estrellas)

Especifique el URL del conjunto de datos. Si es el caso de 4 y 5 estrellas, especifique el nombre del conjunto de datos en la URI de acceso, de modo que el recurso pueda ser identificado y ubicado fácilmente.

Ejemplo de URI: usando una hoja de cálculo

Retomando el ejemplo ya utilizado previamente (Ver: Convertir el formato del archivo a un formato abierto) que contiene información relacionada con el censo de hogares en formato Excel, como se muestra en la siguiente imagen:

	A	B	C	D	E	F	G	H	I	J	K	L
1	CUADRO 1											
2	Población por departamento de residencia habitual, según área y sexo. Total País											
3												
4												
5	Departamento	Total	Hombres	Mujeres	Ignorado	Área Urbana						
6					(1)	Total	Hombres	Mujeres	Ignorado			
7												
8	Total	3.286.314	1.577.725	1.708.481	108	3.110.701	1.478.967	1.631.626	108			
9												
10	Montevideo	1.319.108	613.990	705.014	104	1.305.082	604.965	700.013	104			
11												
12	Resto País	1.967.206	963.735	1.003.467	4	1.805.619	874.002	931.613	4			
13												

Y que luego le quitamos filas de títulos y resúmenes:

Departamento	Total	Hombres	Mujeres	Ignorado	Total Urbano	Hombres Urbano	Mujeres Urbano	Ignorado
Total	3.286.314	1.577.725	1.708.481	108	3.110.701	1.478.967	1.631.626	108
Montevideo	1.319.108	613.990	705.014	104	1.305.082	604.965	700.013	104
Resto País	1.967.206	963.735	1.003.467	4	1.805.619	874.002	931.613	4
Artigas	73.378	36.079	37.299	-	69.854	34.102	35.752	-

Podemos en este caso, usar lo que era el título de la hoja de cálculo como parte de la URI y de esa manera será posible identificar fácilmente el archivo.

http://www.miservidordedatos.com/resultadosfinales/cuadros/NIVEL_NACIONAL/POBLACION_POR_DEPARTAMENTO_RESIDENCIA/segun_area_ysexo.zip

3. Publicación de los conjuntos de datos en el sitio Web institucional

Una vez que se han preparado los conjuntos de datos (datasets), seleccionando y priorizando adecuadamente, utilizando las mejores prácticas en el arreglo de los datos y respondiendo las necesidades de los grupos de interés y ciudadanía en general, ponga sus datos abiertos en línea.

Ubique la información liberada en la sección de datos abiertos del sitio Web de su institución, o en la sección donde expone la información para el público en general. El objetivo es reemplazar la información de formato tradicional a datos abiertos, sumando nuevos conjuntos de datos, siguiendo el programa de liberación de datos. Es importante recordar, si la institución planea en primera instancia liberar información utilizando archivos que ya están expuestos en el sitio Web en formato PDF, eliminar cualquier contraseña y restricción impuesta para uso del archivo (de impresión, copia de texto o descarga). Esto la colocará en nivel de 1 estrella.

Permita que sus datos puedan ser descargados masivamente, en series enteras de tiempo u otro criterio, y si es posible, también a través de APIs; coloque URI a sus conjuntos de datos publicados de 4 y 5 estrellas.

4. Publicación de los conjuntos de datos en el Portal Nacional datosabiertos.gob.pa

Una vez que los datos abiertos han sido publicados en el sitio Web de su institución se debe entrar al portal nacional (datosabiertos.gob.pa) para cargar la información referente a estos datos.

Para el ingreso al portal nacional, a cada institución se le facilitará usuario y contraseña de ingreso a la plataforma, a través de la Autoridad de Innovación AIG. La plataforma sobre la cual está el portal nacional es conocida como “CKAN” diseñada por [Open Knowledge Foundation \(OKF\)](http://openknowledgefoundation.org). CKAN está orientada a proveedores de datos siendo una solución completa de software para hacer los datos accesibles y reutilizables al proveer herramientas para publicar, compartir, encontrar y usar los datos (incluyendo almacenamiento de datos y provisión de APIs de datos robustas). Esta plataforma es muy popular y utilizada en muchas organizaciones de distintos países.

Al cargar los datos abiertos de su institución en el portal nacional, se genera el catálogo nacional de datos abiertos, que no es otra cosa que la lista de todos los conjuntos de datos cargados por todas las instituciones, con sus metadatos, licencias, datos y apps disponibles.

Para mayor claridad sobre lo que representa el sistema de datos abiertos, presentamos la pantalla de creación de un conjunto de datos en el mismo:

Título: ej. Un título descriptivo

* URL: test.catalogodatos.gub.uy/dataset/<dataset> [Edit](#)

Frecuencia de Actualización:

Sistema de Referencia: ej. SIRGAS-ROU98

Cobertura Espacial: ej. Uruguay

Cobertura Temporal: ej. 2013

Descripción: ej. Algunas notas útiles sobre los datos

You can use [Markdown formatting](#) here

Etiquetas: ej. economía, salud mental, gobierno

Licencia:

 Definiciones de licencias e información adicional puede ser encontrada en opendefinition.org

Organismo:

Visibilidad:

Fuente:

Versión:

Autor:

Email del Autor:

Mantenedor:

Email del Mantenedor:

La pantalla “Agregar datos” cuando se trata de la carga de un fichero directamente al portal:

1 Crear conjunto de datos 2 Agregar datos

Fichero:

Nombre:

Descripción:
You can use [Markdown formatting](#) here

Formato:

ⓘ This will be guessed automatically. Leave blank if you wish

La pantalla de mantenimiento de los conjuntos de datos:

🏠 / Pizarra

Feed de Noticias Mis Conjuntos de Datos Mis organismos Mis Categorías

Feed de Noticias Actividad de los elementos que Usted sigue Actividad de: Todo ▾

- Administrador Catalogo actualizó el application Montevideo en Bici del conjunto de datos Bicicircuitos, bicicetarios, estaciones y talleres de reparación hace 3 días
- Administrador Catalogo eliminó el elemento relacionado GXbus hace 3 días
- Administrador Catalogo eliminó el elemento relacionado Bus UY hace 3 días
- Administrador Catalogo actualizó el application Accede Fácil del conjunto de datos Espacios accesibles de Montevideo hace 3 días

5. Difusión y promoción de los datos abiertos

La difusión y promoción del uso de los datos abiertos se inicia con las herramientas que ya existen en la institución: el sitio Web (página de datos abiertos y sección de noticias, por ejemplo) y el correo electrónico. Es importante que cada vez que sean cargados nuevos conjuntos de datos abiertos se promocionen en ambas instancias. Otras herramientas importantes para difundir la apertura de datos son las redes sociales las cuales pueden ser utilizadas para envío de alertas de liberación de nuevos conjuntos de datos a usuarios seguidores, además de promocionar su uso y re-uso. Las redes sociales son muy útiles para publicar información de forma rápida, y llevar mensajes cortos y puntuales sobre datos publicados recientemente.

Al utilizar el correo electrónico como medio de difusión es recomendable identificar previamente a usuarios y sus áreas de interés, lo que posibilitará estructurar listas de acuerdo a los perfiles de interés de estos. Sin embargo, la promoción a través de estas listas requiere de una planeación e investigación en la comunidad, instituciones de gobierno y empresa privada, de tal forma, que se pueda conformar efectivamente las mismas. Los mensajes en los correos deben ser explicativos e incluir contenidos sobre los conjuntos de datos liberados.

Para la obtención de los mejores resultados en la difusión y promoción del uso y re-uso de los datos abiertos, es necesario realizar una planificación y programación de actividades incluyendo tanto a los grupos de interés y comunidad en general, como a las unidades generadoras de los datos abiertos en las instituciones, en la medida de lo posible. Por ejemplo, al realizar actividades tipo paneles de discusión se facilita la interacción entre usuarios y generadores de datos abiertos con el objeto de intercambiar ideas, opiniones y sugerencias, lo que redundará en el mejoramiento de los servicios de información por parte de las instituciones de gobierno, tomando en cuenta de manera directa la opinión de la comunidad.

Otros métodos que pueden ser utilizados para generar y mantener el interés de los grupos y comunidad en general en el uso y re-uso de los datos abiertos de gobierno, es publicar boletines con la descripción de los datos que serán liberados de acuerdo a la programación institucional de apertura de datos. Los eventos especiales y actividades con la comunidad son ocasiones apropiadas para promocionar los datos abiertos. En ellos se pueden hacer presentaciones sobre uso y re-uso de los datos abiertos, como visualizaciones de experiencias en otros países y enseñarles el valor agregado que se puede sacar de ellos diseñando aplicaciones en temas especializados.

GLOSARIO DE TÉRMINOS

ANONIMIZACIÓN - El proceso de adaptación de datos de manera que los individuos no pueden ser identificados a partir de ella.

API- (*Application Programming Interface*) Una manera en que los programas informáticos se hablan entre ellos. Puede ser entendido en términos de cómo un programador envía instrucciones entre los programas.

CSV – Formato de archivos separados por coma, donde cada valor es delimitado por una coma o punto y coma; es una forma sencilla de publicar información pero necesariamente se debe publicar también información sobre los datos que contiene (meta data). De otra manera esta información será difícilmente interpretada.

DBF – Formato de archivos utilizado por dBase, generalmente es el formato del archivo de base de datos para computadores personales.

DBMS – Database Management System; sistema de gestión de bases de datos.

DOC – Formato de archivo creado por la herramienta de procesamiento de textos de Microsoft.

DTD – Describe el formato de los datos (Data Type Definition) en archivos XML, HTML, etc.

INE – Instituto Nacional de Estadísticas.

JSON – JavaScript Object Notation es un lenguaje muy sencillo de automatizar su procesamiento. Si bien no está tan expandido como XML su filosofía similar y su fácil interpretación lo hacen un lenguaje muy apropiado para el intercambio de información.

KML – Keyhole Markup Language, formato de archivo utilizado para la representación de datos geográficos en tres dimensiones, basado en XML.

LINKED DATA (Datos Relacionados) – Forma parte de la web semántica, a diferencia de la web actual donde los documentos de hipertexto se relacionan a diferentes puntos en el HTML, esta nueva forma enlaza los datos con cosas arbitrarias que se describen en formato RDF.

METADATOS – Datos que describen otros datos.

ODS – Formato de archivo de Open Document, es un estándar abierto para documentos de hojas de cálculo.

ONTOLOGÍA – En ciencias de la información y ciencias de la computación se entiende por ontología a la representación del conocimiento como un conjunto de conceptos pertenecientes al mismo dominio y las relaciones entre estos conceptos.

OWL – Es una recomendación de W3C y fue diseñado para proveer un formato común para el procesamiento de la información en la web, es un lenguaje para ser procesado por computadoras y no está diseñado para mostrar información a los usuarios. Es parte de la web semántica.

PDF – Formato portable de documentos, estándar para la publicación de documentos en la web.

RDF – Es un marco para la descripción de recursos en la web. El formato RDF permite mediante un documento XML describir recursos publicados en Internet.

SHP – Formato de archivo informático de datos espaciales desarrollado por la empresa ESRI comercializadora de productos ArcGIS y ArcInfo.

SPARQL – Motor de consultas para la web semántica.

TXT – Formato de archivo de texto plano.

URI – Identificador único de recurso, permite identificar únicamente un recurso, pagina, documento, dirección de correo electrónico, etc.

VOCABULARIOS– Es un conjunto de términos acordados y agrupados en un espacio de nombres que sirven para describir cosas.

XLS – Formato de archivo creado por Microsoft para su herramienta de hoja de cálculo.

XML – Formato extensible de marcas, lenguaje que permite la definición de etiquetas o marcas derivado del SGML.

W3C – Consorcio internacional que produce recomendaciones para la WWW.

WEB SEMÁNTICA – La web semántica es una visión futura de la web donde la información publicada debe tener un significado exacto, esta información debe ser entendida y procesada por computadoras. “Es una web de datos que pueden ser procesados directa o indirectamente por maquinas” según la definición de Tim Berners-Lee.

WEBSERVICES – En el caso que la información sea extraída de una fuente de información viva, en constante modificación; se recomienda la utilización de servicios web que expongan dicha información en tiempo real.