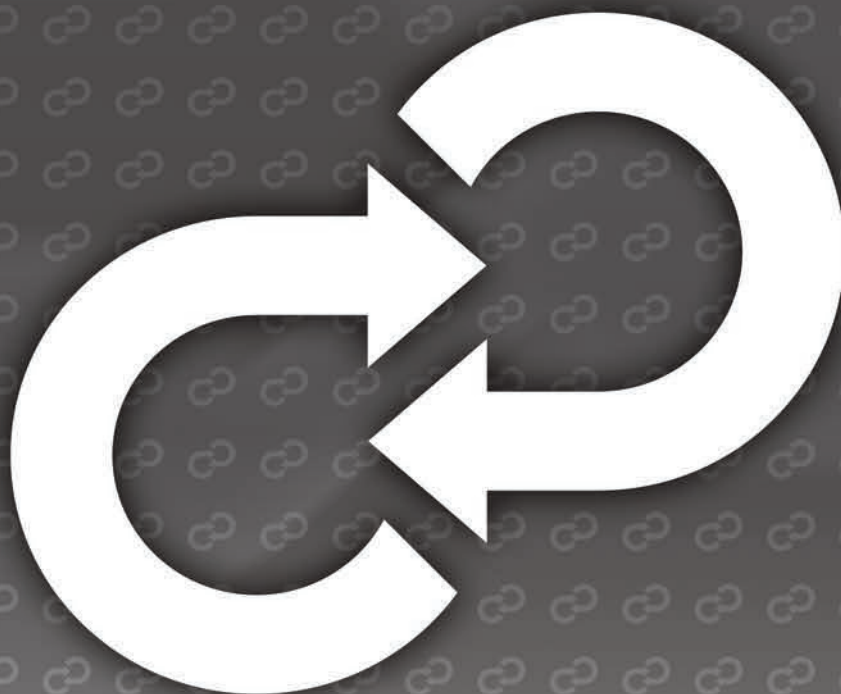


# Guía básica de apertura y de reutilización de datos abiertos de gobierno



GUIA BASICA DE APERTURA Y DE  
REUTILIZACIÓN

# Datos Abiertos de Gobierno

BUENAS PRÁCTICAS

Versión 1.0 – 2012

*Este documento ha sido elaborado por AGESIC (Agencia para el Desarrollo del Gobierno de Gestión Electrónica y la Sociedad de la Información y el Conocimiento)*

*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.*





### ***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 profundizar en el tema.*



Capítulo I

# Introducción



## Introducción

### ¿Qué son los Datos Abiertos?

Datos Abiertos son aquellos datos que están disponibles libremente para su utilización, reutilización y redistribución.

Los Datos Abiertos deben cumplir con las siguientes premisas.

- **Disponibilidad y acceso:** Los datos deben estar disponibles un costo razonable y de una forma conveniente y modificable.
- **Reutilización y redistribución:** El formato de los datos permitirá su reutilización, redistribución e integración a otros conjuntos de datos.
- **Participación universal:** Todas las personas deben poder utilizar, reutilizar y redistribuir la información sin restricciones.

En diciembre de 2007 la organización Opengovdata publicó ocho principios<sup>1</sup> que deben ser considerados por los gobiernos a la hora de publicar datos. Estos principios no indican cuáles datos deberán ser públicos, sino que especifican las condiciones para que los datos públicos sean, además, abiertos.

#### 1. Los datos deben ser completos

Todos los datos estarán disponibles. Los datos se almacenan electrónicamente incluyendo documentos, bases de datos, transcripciones, grabaciones de audio y video, imágenes, etc. Los datos públicos son datos que no están sujetos a las limitaciones vigentes de privacidad, seguridad o privilegio, que se rige por otras leyes.

#### 2. Los datos deben ser primarios

Los datos se publican tal como se extraen de la fuente de origen, con el mejor nivel posible de granularidad, no en formas agregadas o modificadas.

#### 3. Los datos deben ser oportunos

Los datos se hacen disponibles tan pronto como sea necesario para preservar el valor de los mismos.

---

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



**4. Los datos deben ser accesibles**

Los datos están disponibles para la gama más amplia de usuarios y para la gama más amplia de propósitos.

**5. Los datos deben ser procesables por máquina**

Los datos estarán razonablemente estructurados para permitir el procesamiento automático de los mismos.

**6. El acceso no debe ser discriminatorio**

Los datos están disponibles para cualquier persona, sin necesidad de registro.

**7. El formato de los datos no debe ser propietario**

Los datos están disponibles en un formato sobre el cual ninguna entidad tiene control exclusivo.

**8. Los datos deben ser libres de licencia**

Los datos no están sujetos a ningún derecho de autor, patentes, marcas o reglamento de secreto comercial. Restricciones razonables de privacidad, seguridad y de privilegios se pueden permitir ya que se rigen por otras leyes.

## ¿Qué datos de gobierno podrían ser abiertos?

Los datos públicos de gobierno según se definen en la Ley N° 18.381 de Acceso a la Información Pública. El punto clave a destacar es que la apertura de datos se centra en los datos no personales que no están sujetos a restricciones específicas.

Capítulo II

# Apertura de Datos



# Apertura de Datos

## Clasificación de los datos abiertos

Tim Berners-Lee<sup>2</sup> propuso una categorización del grado de apertura de los datos basado en estrellas, respecto 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.

**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.

### 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.

En el siguiente capítulo se explica 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 disponibilizados en un sitio Web.

Entonces, a modo de resumen:

Cualquier archivo disponible en sitios Web en formato .PDF, .DOC, .ODT, .TXT, entre otros; y que esté bajo una licencia abierta será clasificado como dato abierto de 1 estrella.

---

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

## 2 Estrellas

El siguiente nivel de clasificación, es 2 estrellas. 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 máquinas.

Por ejemplo, datos que se encuentren en formato .XLS y organizados en una tabla se podrían clasificar como 2 estrellas.

## 3 Estrellas

El siguiente nivel de clasificación, 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 podrían clasificar como 3 estrellas. Otro ejemplo es un archivo en formato .XML.

## 4 y 5 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 pueden ser RDF y SPARQL.

Los niveles 4 y 5 son considerados avanzados y se acercan mucho al ideal de publicación de un dato abierto. Dado que en la actualidad, en Uruguay se está comenzando con la apertura de datos, se aspira en estas primeras que todos los conjuntos de datos abiertos de gobierno alcancen como mínimo el nivel 3.

No obstante, se espera ir mejorando de manera gradual y planificada para alcanzar los niveles máximos en los casos de conjuntos de datos que amerite el esfuerzo.

## 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**

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

[http://semanticweb.org/wiki/Main\\_Page](http://semanticweb.org/wiki/Main_Page)

- **Para desarrolladores PHP**

Graphite es una librería para PHP construida sobre ARC2 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/linqtordf/>

- **Implementaciones SPARQL**

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

- **Protégé**

Es un editor de ontologías gratuito y de código abierto.

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

## Procedimiento para la apertura de datos

### Reglas básicas para la apertura de datos

El manual de datos abiertos (Open Data HandBook<sup>3</sup>) recomienda seguir tres reglas básicas en el proceso de publicar información de datos abiertos.

Mantenerlo simple. Como primer paso a dar, hay que abrir conjuntos de datos pequeños, sencillos y lo más rápido posible; incluso puede ser una parte de un conjunto de datos más amplio.

Comprometerse temprano y comprometerse con frecuencia con los potenciales usuarios y re utilizadores de los datos; sean estos ciudadanos comunes, empresas o desarrolladores. Hay que tener en cuenta también que la mayor parte de los datos no le llegarán al usuario común sino a un intermediario (empresa o desarrollador) que va a procesar dicha información y la mostrará en un formato más amigable. Por lo tanto si van a existir aplicaciones que dependan de la información publicada es importante que la información sea oportuna y este siempre disponible.

Aclarar miedos y dudas en las etapas tempranas. Cuando se va a abrir datos es lógico que surjan dudas, malentendidos y miedos, es importante identificar cuáles son los más importantes y direccionarlos a quien pueda aclarar los problemas que se plantean de manera de poder resolverlos en las primeras etapas del proceso.

En una primera instancia, la forma más sencilla de comenzar con la apertura de datos se debe tratar de elegir información que no insuma demasiados recursos en su elaboración, reutilizar información que esté siendo distribuida en otros formatos y hacerla disponible en formatos abiertos. Se debe tener en cuenta también al seleccionar un conjunto de datos que a veces la información que para una organización es muy importante no lo es para los reutilizadores que son quienes harán uso de la misma en sus aplicaciones o para los posibles consumidores. Por lo tanto el ejercicio de comparar con otras experiencias es un insumo a tener en cuenta llegado el momento de elegir que datos abrir. Con el conjunto de datos elegido se debe determinar si existen derechos de propiedad intelectual; en base a esto se debe aplicar una licencia<sup>4</sup> abierta que contemple estos derechos y soporte los conceptos de datos abiertos vistos anteriormente.

---

3 <http://opendatahandbook.org>

4 <http://opendefinition.org/licenses/>

El paso siguiente es dejar los datos disponibles en un formato reutilizable y en su forma más pura.

## Primeros pasos a seguir

A continuación se describe un conjunto de pasos recomendados para quien quiere abrir datos y no sabe por donde empezar:

1. Seleccionar el conjunto de datos a abrir
2. Convertir el formato del archivo a un formato abierto
3. Aplicar la licencia abierta
4. Incluir los metadatos
5. Empaquetar
6. Asignar URI

**Nota:** Por más información de cómo publicar en el Catálogo Nacional de Datos Abiertos, [catalogodatos.gub.uy](http://catalogodatos.gub.uy), consulte la” Guía Rápida de Publicación en [datos.gub.uy](http://datos.gub.uy)”

### 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 abrir.

Lo principal para comenzar es identificar los datos que tenemos publicados o que son de fácil transformación. Para comenzar este proceso puede realizarse las siguientes preguntas:

- ¿Qué información le solicitan habitualmente y que exige un procesamiento de datos?
- ¿Qué información existe en su oficina con datos que pueden ser de interés para la población?
- ¿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 a ser publicada, debe seleccionar la que se encuentre en condiciones de ser publicada. Tenga en cuenta aspectos legales, de completitud, capacidad para mantenerla actualizada, formato, etc.

## Convertir el formato del archivo a un formato abierto

En este apartado mostraremos 2 posibles casos de transformar 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

#### Ejemplo de .XLS a .CSV

A modo de ejemplo tomamos información relacionada con el censo de hogares en formato Excel y realizaremos la transformación a formato abierto CSV.

Departamento	Total	Hombres	Mujeres	Ignorado (1)	Área Urbana			
					Total	Hombres	Mujeres	Ignorado (1)
<b>Total</b>	<b>3.286.314</b>	<b>1.577.725</b>	<b>1.708.481</b>	<b>108</b>	<b>3.110.701</b>	<b>1.478.967</b>	<b>1.631.626</b>	<b>108</b>
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

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.

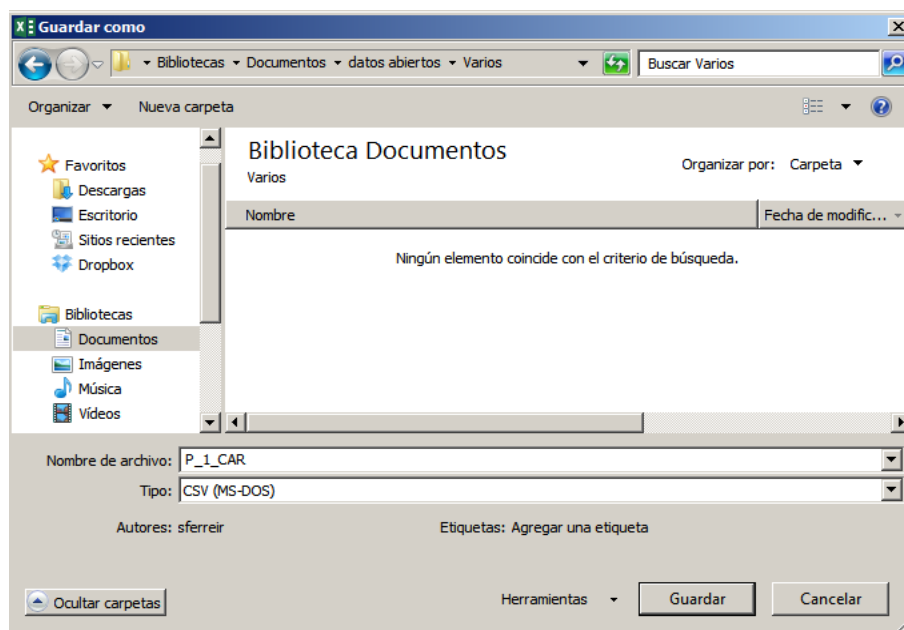
Como primer paso quitamos información adicional dejando solamente información de nombres de columnas.

Departamento	Total	Hombres	Mujeres	Ignorado	Total Urbano	Hombres Urbano	Mujeres Urbano	Ignorado
<b>Total</b>	<b>3.286.314</b>	<b>1.577.725</b>	<b>1.708.481</b>	<b>108</b>	<b>3.110.701</b>	<b>1.478.967</b>	<b>1.631.626</b>	<b>108</b>
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, archivo separado por comas y si 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 Pa;s;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, la extracción de datos a formato XML.

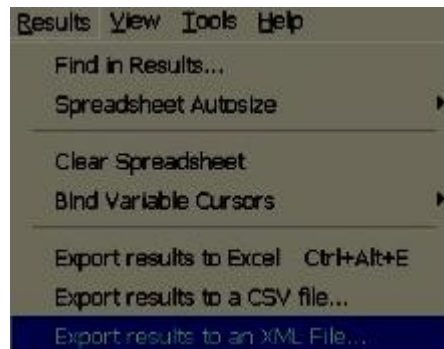


### Ejemplo: exportación de datos en formato XML en Oracle

---

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.

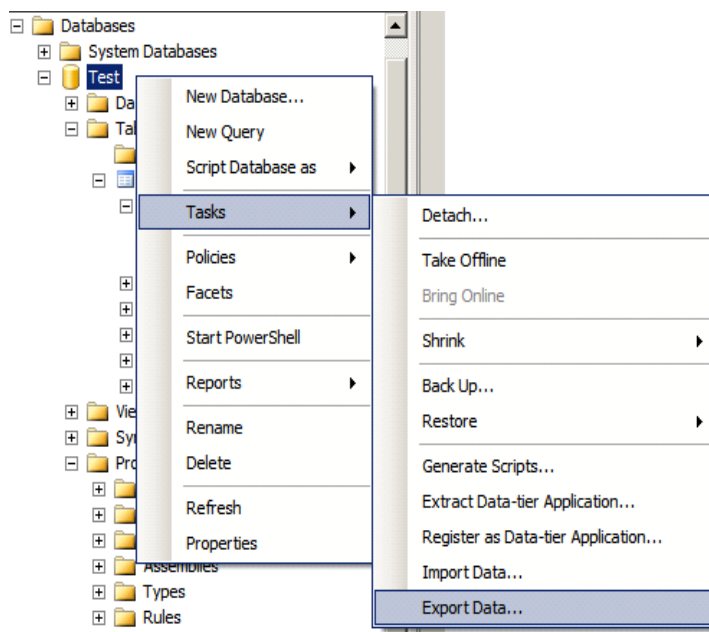
Sin embargo existen otras herramientas que permiten realizar la exportación, por ejemplo PL/SQL Developer permite realizar la exportación directa de datos a formato XML y CSV desde un menú como muestra la siguiente imagen:



### Ejemplo: exportación de datos en formato XML en 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 en formato XML en 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:

idalgo	algo
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="idalgo" Type="int(11)" Null="NO" Key="PRI"
Extra="" Comment="" />
<field Field="algo" Type="varchar(50)" Null="NO" Key="" Extra=""
Comment="" />
<key Table="test" Non_unique="0" Key_name="PRIMARY"
Seq_in_index="1" Column_name="idalgo" 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="idalgo">1</field>
<field name="algo">registro 1</field>
</row>
<row>
<field name="idalgo">2</field>
<field name="algo">registro 2</field>
</row>
<row>
<field name="idalgo">3</field>
<field name="algo">registro 3</field>
</row>
<row>
```

```
<field name="idalgo">4</field>
<field name="algo">registro 4</field>
</row>
</table_data>
</database>
</mysqldump>
```

## Aplicar la licencia abierta

Aplicar una licencia de datos abiertos es esencial para que el dato publicado sea clasificado como abierto. 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 disponibilizados en un sitio Web.

Parte del texto que podría contener una licencia abierta es el siguiente:

**“El Usuario reutilizador de los datos de este sitio, deberá cumplir, al menos, las siguientes condiciones básicas:**

- **Mantener el sentido original de la información**
- **Siempre citar la fuente**
- **Explicitar la fecha de la última actualización.**

**Se permite cualquier explotación de los datos abiertos, incluyendo una finalidad comercial, así como la creación de obras derivadas, estando permitida su distribución sin ninguna restricción.”**

Entonces, a modo de resumen, siempre que se publiquen datos para considerarlos abiertos, se debe especificar la licencia.



### **Ejemplo: ¿cómo se indica la licencia en el Catálogo Nacional de Datos Abiertos?**

A la hora de catalogar se ofrece la opción de utilizar la licencia para datos abiertos de gobierno de Uruguay que se encuentra publicada en el sitio <http://datos.gub.uy>.

También existe la posibilidad de utilizar otra licencia, y en ese caso, el usuario Catalogador deberá acompañar al archivo de datos abiertos con otro archivo que contenga el texto de la licencia.

## Incluir los metadatos

Para facilitar la utilización de los datos por parte de terceros se les debe de proveer de información adicional (metadatos) que describan la información que se publica. Para el caso de estructuras relacionales u hojas de cálculo puede ser la descripción de las columnas, para archivos XML puede ser el DTD, etc.



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

A la hora de disponibilizar los datos abiertos en un formato .XLS o .ODS o .CSV, además se debe publicar otro con la licencia (en el caso que aplique) y un tercer archivo con los metadatos.

Este archivo debe explicar el significado de cada columna, como se muestra en la siguiente imagen:

Nombre de atributo	Descripción	Tipo de dato	Recurso relacionado	Información Adicional
<p>&lt;Descripción exacta de la columna tal cual aparece en los archivos de los conjuntos de datos&gt;. <b>Ejemplo:</b> <b>CodigoDepto</b></p>	<p>&lt;Descripción detallada del significado del atributo orientada a las personas que pretendan utilizar el atributo&gt;.  <b>Ejemplo:</b> <b>Departamento de Residencia</b></p>	<p>&lt;Numérico, Texto, Fecha/Hora, Moneda&gt; <b>Ejemplo:Numérico</b></p>	<p>&lt;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.&gt;  <b>Ejemplo:Departamentos.xls</b></p>	<p>&lt;Cualquier información adicional sobre el atributo que complemente la información básica.&gt; <b>Ejemplo:</b>En el archivo <i>Departamento.xls</i> están los nombres de los departamentos asociados a dicho código.  <b>Ejemplo:</b> <b>1- Montevideo;</b> <b>2-Soriano;</b> <b>3-Paysandu....</b></p>

## 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 correspondientes, se debe empaquetar toda la información utilizando herramientas para la compresión de archivos.

## Asignar URI

Es recomendable realizar cambios en la URI de acceso de modo que el recurso pueda ser identificado por su nombre fácilmente.

### Ejemplo: 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:

Departamento	Total	Hombres	Mujeres	Ignorado (1)	Área Urbana			
					Total	Hombres	Mujeres	Ignorado (1)
<b>Total</b>	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

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





Capítulo III

# Reutilización de Datos



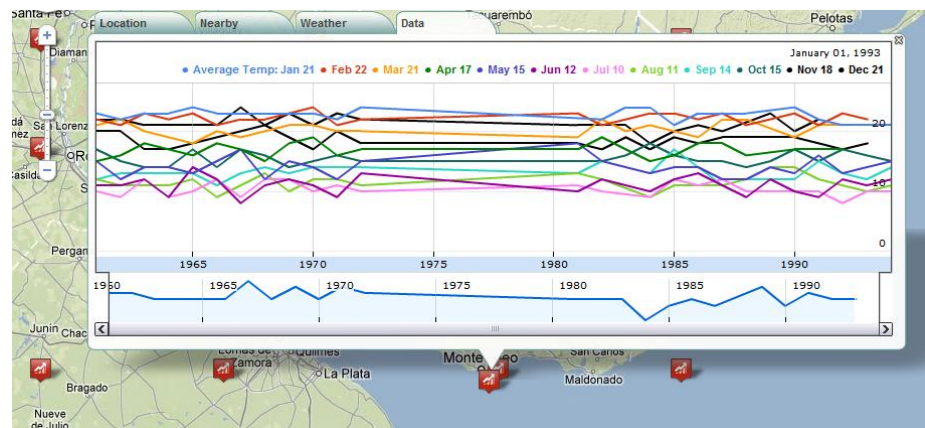
## Reutilización de datos

### Ejemplos de reutilización de datos

#### GEOME

Es un sitio que hace uso de datos abiertos para generar información<sup>5</sup> de diferentes tipos.

Algunas de las aplicaciones presentan Indicadores del Banco Mundial, información sobre las elecciones en UK en el año 2010 y mapa de clima mundial. Este último basado en información de más de mil quinientas estaciones meteorológicas del mundo recopilada por la Organización mundial de meteorología<sup>6</sup>.






Fuente: [http://labs.geo.me/climate\\_data](http://labs.geo.me/climate_data)

<sup>5</sup> [www.geo.me](http://www.geo.me)


<sup>6</sup> [http://www.wmo.int/pages/index\\_en.html](http://www.wmo.int/pages/index_en.html)

## Ottawa

Otro caso de uso de datos abiertos es el de la ciudad de Ottawa, donde existe publicado un catálogo de datos en su sitio institucional <http://ottawa.ca> el cual utilizaremos para mostrar un ejemplo práctico de utilización de este tipo de información en formato abierto.

M	Name	CSV/XLS	XML	DWG	KML	SHP	RSS
	Murals – Paint it Up City Locations						
	Museums 2010						

[ Top ]

N	Name	CSV/XLS	XML	DWG	KML	SHP	RSS
	Needle Drop Boxes and Take it Back! pharmacies						

Fuente: [http://ottawa.ca/en/city\\_hall/statisticsdata/opendata/info/index.htm](http://ottawa.ca/en/city_hall/statisticsdata/opendata/info/index.htm)

El catálogo de datos muestra ordenado alfabéticamente los conjuntos de datos disponibles y el formato de los archivos que contiene.

Descargando el archivo KML correspondiente a “Museums 2010”, podemos observar cómo se estructuran los datos para poder ser visualizados en la interface de mapas de Google Maps. A modo de ejemplo se muestra un fragmento de esta información donde en la primer parte se ve información general de diseño, como son imágenes y colores, iconos y su tamaños, etc. a utilizar en los mapas.

```
<?xml version="1.0" encoding="UTF-8"?>
<kml xmlns="http://www.opengis.net/kml/2.2"
xmlns:gx="http://www.google.com/kml/ext/2.2"
xmlns:kml="http://www.opengis.net/kml/2.2"
xmlns:atom="http://www.w3.org/2005/Atom">
<Document>
  <name><![CDATA[2010_Museums]]></name>
  <open>1</open>
  <Snippet maxLines="0"><![CDATA[]]></Snippet>
  <description><![CDATA[Exported from 2010_Museums on 2010-Jun-17]]></description>
  <Style id="FEATURES">
    <IconStyle>
      <color>FF993300</color>
      <scale>0.3333333333333333</scale>
    <Icon>
      <href>root://icons/palette-4.png</href>
      <x>32</x>
      <y>128</y>
      <w>32</w>
      <h>32</h>
```

```

        </Icon>
    </IconStyle>
    <LabelStyle>
        <color>00FFFFFF</color>
    </LabelStyle>
    <BalloonStyle>
        <text>
        </text>
    </BalloonStyle>
</Style>
.....
    
```

En este archivo también se especifican las coordenadas, altura, etc. para los lugares que se desean ubicar en el mapa. En nuestro ejemplo utilizamos el museo Ottawa Room.

```

<Placemark>
    <name><![CDATA[Ottawa Room]]></name>
    <Snippet maxLines="0"><![CDATA[]]></Snippet>
    <styleUrl>#FEATURES</styleUrl>
    <Point>
        <extrude>0</extrude>
        <altitudeMode>clampedToGround</altitudeMode>
        <coordinates>
            -75.6955319434698,45.4202457646201,0
        </coordinates>
    </Point>
</Placemark>
    
```

El mismo conjunto de datos permite descargar el archivo SHP que contiene información adicional en un formato .DBF que si lo abrimos con una hoja de cálculo vemos información de los lugares contenidos en el archivo KML.

NAME, C, 150	ADDRESS, C, 75	TELEPHONE, T, 15	TYPE, C, 60	CITYNAME, FR, C, 50
Billings Estate National Historic Site	2100 Cabot Street, Ottawa	613-247-4830	Museum	18 Lieu historique national du domaine Billings
Cumberland Heritage Village Museum	2940 Old Montreal Road, Cumberland	613-833-3059	Museum	19 Musée-village du patrimoine de Cumberland
Goulbourn Museum	2064 Huntley Road, Stittsville	613-831-2393	Community Museum	21 Musée de Goulbourn
Ottawa Jewish Archives	21 Nadolny Sachs Private, Ottawa	613-798-4696	Community Museum	15 Archives juives d'Ottawa
Nepean Museum	16 Rowley Avenue, Nepean	613-723-7936	Community Museum	8 Musée de Nepean
Osgoode Township Historical Society and Museum	7814 Lawrence Street, P.O. Box 74, Vemon	613-821-4062	Community Museum	20 Musée et société historique du canton d'Osgoode

Uniendo estas fuentes de información en una aplicación podemos encontrar fácilmente la ubicación de los museos de la ciudad en un mapa, conocer su dirección exacta y horarios de funcionamiento, etc.

La aplicación desarrollada para la ciudad de Ottawa no solo contiene información sobre museos, también se puede encontrar otros sitios de interés como parques, centro recreativos y deportivos.

En la siguiente imagen se puede apreciar cómo se visualiza parte de la información analizada en la aplicación <http://www.ottawafun.ca/>; donde además de tener los puntos ubicados en el mapa, también se puede filtrar la información por orden alfabético y por los mejor puntuados.

Al posicionar el puntero del mouse en un punto se accede también información adicional.



Fuente: <http://www.ottawafun.ca/>

En Uruguay como ejemplo podemos tomar lo publicado por la Intendencia de Montevideo y la aplicación. GXBus.

## GXBus

Es una aplicación nativa para dispositivos móviles que permite encontrar recorridos y horarios de ómnibus para llevar al usuario de un punto a otro de Montevideo. Fue desarrollada utilizando GeneXus X Evolution 2 y es la ganadora del Smart Challenge 2011 de Artech en el marco del XXI Encuentro GeneXus.



Fuente: [www.gxbus.com](http://www.gxbus.com)



## Cómo ir

Aplicación para obtener rutas en ómnibus (incluyendo horarios) o a pie entre dos puntos de la ciudad.

### Cómo ir

#### Origen y destino

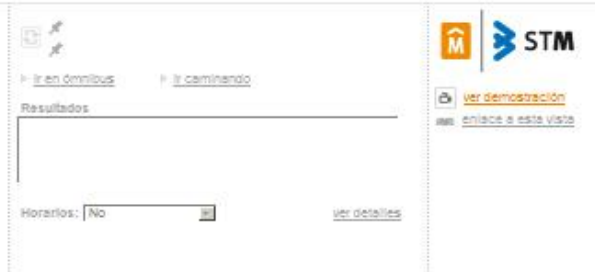
Tipo de ubicación:

Calle:

Esquina / Número:

[marcar origen](#)


[marcar destino](#)

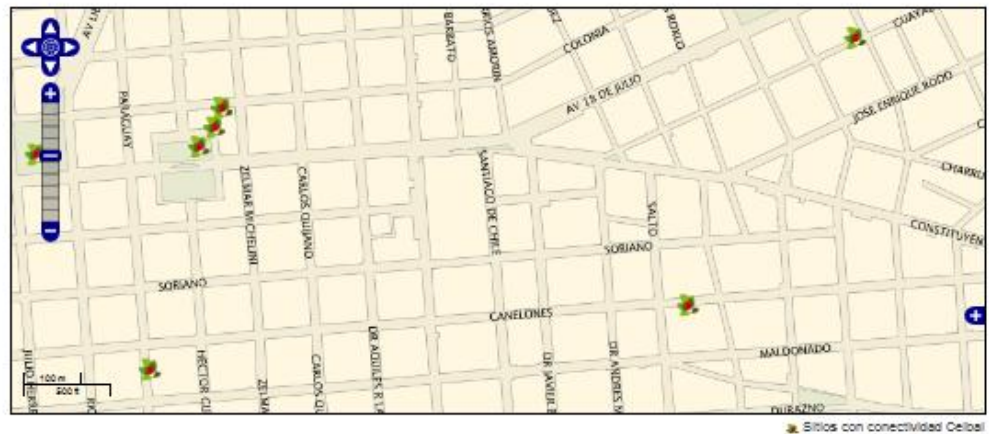


ir en ómnibus     caminando

Resultados

Horarios:  [ver detalles](#)


  
[ver demostración](#)
  
[enlace a esta vista](#)



Fuente: [www.montevideo.gub.uy/aplicacion/como-ir](http://www.montevideo.gub.uy/aplicacion/como-ir)

**Horarios de ómnibus:** Consulta de horarios de ómnibus de transporte metropolitano. A partir del número de línea y el tipo de día (Hábil, Sábado, Domingos y feriados) devuelve la lista de horarios de todas las variantes en los diferentes puntos notables de su recorrido (puntos de control).

Sitio web: [www.montevideo.gub.uy/aplicacion/horarios-de-omnibus](http://www.montevideo.gub.uy/aplicacion/horarios-de-omnibus)

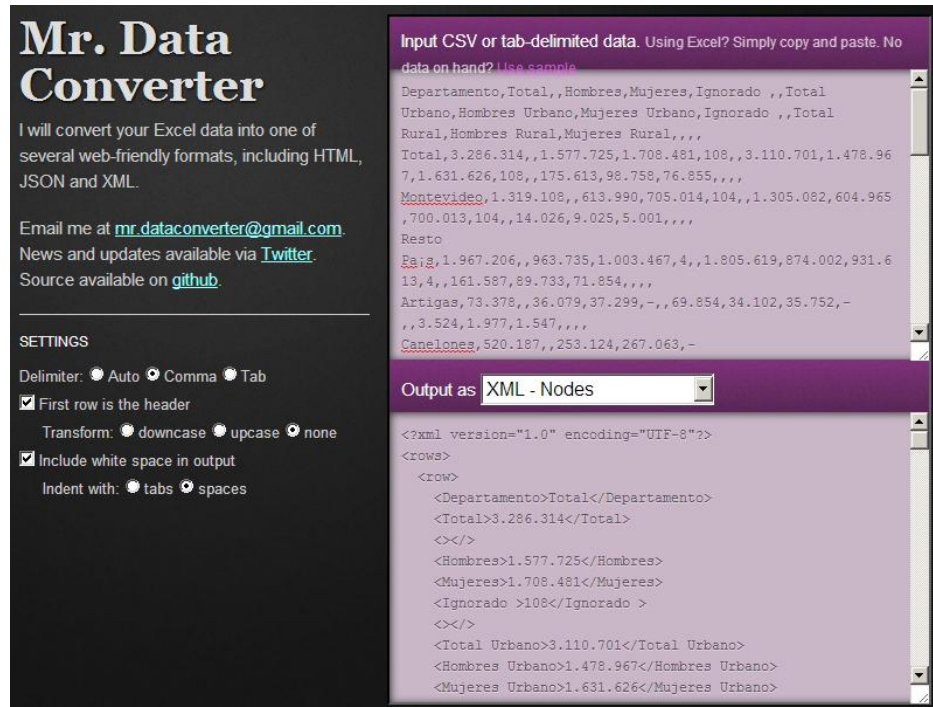
**Open Street Map: OpenStreetMap** es un mapa de todo el mundo que puede ser editado por cualquier persona. Permite ver, editar y usar información geográfica de manera colaborativa desde cualquier lugar del mundo. Parte de la información geográfica publicada por la Intendencia de Montevideo fue utilizada por este sitio para mejorar la información de los mapas de Montevideo. Sitio web: [www.openstreetmap.org/](http://www.openstreetmap.org/)

# Herramientas para el procesamiento de datos

## Herramienta para conversión de datos

Mr. Data Converter es un sitio que permite convertir un archivo de hoja de cálculo o de texto CSV a formato XML, JSON, etc.

La forma de convertir es muy sencilla simplemente se copian los datos en el cuadro superior y se selecciona el formato al cual se quiere convertir; el resultado de la conversión se ve en el cuadro inferior de la pantalla.



**Mr. Data Converter**

I will convert your Excel data into one of several web-friendly formats, including HTML, JSON and XML.

Email me at [mr.dataconverter@gmail.com](mailto:mr.dataconverter@gmail.com).  
 News and updates available via [Twitter](#).  
 Source available on [github](#).

---

**SETTINGS**

Delimiter:  Auto  Comma  Tab

First row is the header

Transform:  downcase  upcase  none

Include white space in output

Indent with:  tabs  spaces

**Input CSV or tab-delimited data.** Using Excel? Simply copy and paste. No data on hand? [Use sample](#)

```
Departamento,Total,,Hombres,Mujeres,Ignorado ,,Total
Urbano,Hombres Urbano,Mujeres Urbano,Ignorado ,,Total
Rural,Hombres Rural,Mujeres Rural,,,,
Total,3.286.314,,1.577.725,1.708.481,108,,3.110.701,1.478.96
7,1.631.626,108,,175.613,98.758,76.855,,,,
Montevideo,1.319.108,,613.990,705.014,104,,1.305.082,604.965
,700.013,104,,14.026,9.025,5.001,,,,
Resto
Eaig,1.967.206,,963.735,1.003.467,4,,1.805.619,874.002,931.6
13,4,,161.587,89.733,71.854,,,,
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,-
```

Output as **XML - Nodes**

```
<?xml version="1.0" encoding="UTF-8"?>
<rows>
  <ROW>
    <Departamento>Total</Departamento>
    <Total>3.286.314</Total>
  </ROW>
  <ROW>
    <Hombres>1.577.725</Hombres>
    <Mujeres>1.708.481</Mujeres>
    <Ignorado >108</Ignorado >
  </ROW>
  <ROW>
    <Total Urbano>3.110.701</Total Urbano>
    <Hombres Urbano>1.478.967</Hombres Urbano>
    <Mujeres Urbano>1.631.626</Mujeres Urbano>
```

## Herramienta para visualización de datos

Junar es una herramienta online que permite el procesamiento de información de manera sencilla para realizar visualizaciones de información en forma de gráficas, datos geo referenciados, etc.

Desde el menú inicial se puede comenzar seleccionando los datos a utilizar. Hay que tener en cuenta que esta herramienta es un servicio web por lo que los datos se deben subir al servidor de Junar para su posterior utilización.



**¡Bienvenido a Junar!**

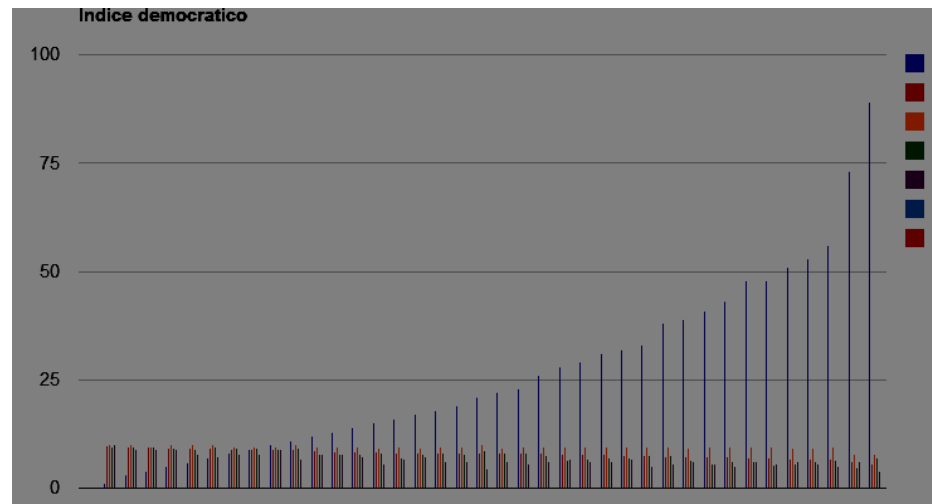
**Abra sus datos de manera sencilla** Siga estos sencillos pasos para empezar.

- Recolecte sus datos**  
 Elija los conjuntos de datos que desea cargar desde el ordenador o la Web.
- Mejora tus datos**  
 Los Conjuntos de datos que has colectado puede ser mejorados en Vistas de datos, Gráficos y Colecciones
- Tu micrositio**  
 Tus datos son automáticamente publicados en tu **Micrositio**, alojado en la plataforma de Junar

Dado el siguiente archivo en formato ODS, veremos el resultado de su procesamiento con Junar.

	Rank	Overall score	Electoral process and pluralism	Functioning of government	Political participation	Political culture	Civil liberties
Norway	1	9,80	10,00	9,64	10,00	9,38	10,00
Denmark	3	9,52	10,00	9,64	8,89	9,38	9,71
Sweden	4	9,50	9,58	9,64	8,89	9,38	10,00
New Zealand	5	9,26	10,00	9,29	8,89	8,13	10,00
Australia	6	9,22	10,00	8,93	7,78	9,38	10,00
Finland	7	9,19	10,00	9,64	7,22	9,38	9,71

El proceso consiste en dos etapas, recolectar y mejorar. En el primer proceso se identifican los datos, se seleccionan y se guardan, en el segundo paso se comienza a trabajar con esta información para mejorar la visualización. El resultado obtenido es un gráfico de barras con la información por países.



Fuente: [www.junar.com](http://www.junar.com)

## Enlaces relacionados

Por mayor información sobre herramientas de extracción de datos, visualizadores de datos, entre otras se ofrecen una serie de enlaces que pueden resultar útiles:

<http://worldbank.tumblr.com/>

<http://www.datatracker.org/category/software/>

<http://red.gnoss.com/comunidad/nextweb>

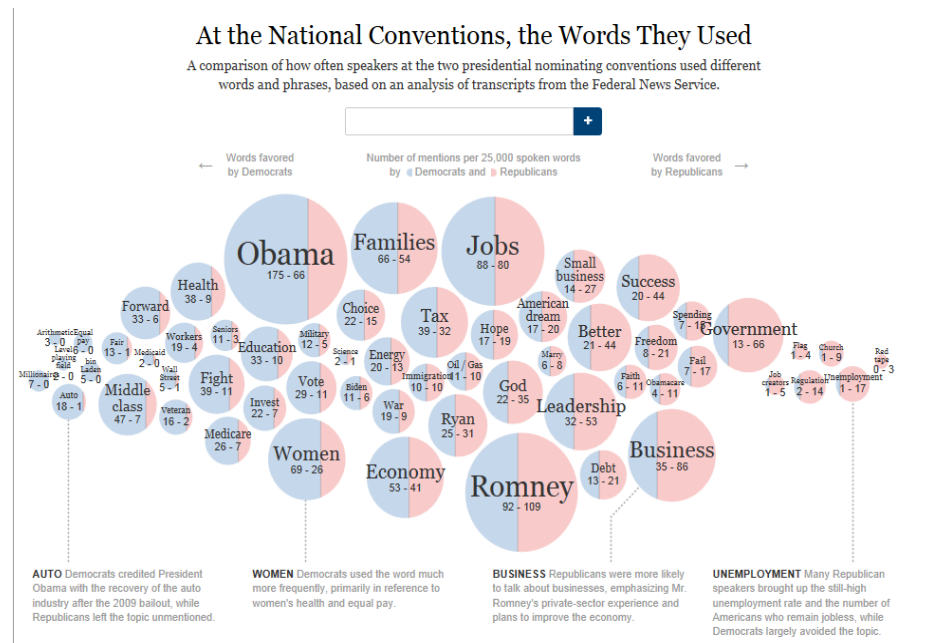
<http://tools.loosewire.org/database-reporting/>

<http://opendata-tools.org/en/>

## Periodismo de datos

### Ejemplos

Otros ejemplos de reutilización son llevados a cabo por periodistas que accediendo a datos abiertos los transforman y enriquecen mediante la utilización de herramientas informáticas. Tal es el caso de las infografías; es una técnica que consiste en mostrar en poco espacio mucha información utilizando imágenes.



En este caso el NY Times<sup>7</sup> publicó las palabras más utilizadas en los discursos de las diferentes convenciones en los Estados Unidos.

### Enlaces a herramientas y más información

Algunas herramientas útiles se pueden encontrar en los siguientes enlaces.

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

<http://www.wolfram.com/cdf/>

<sup>7</sup> [www.nytimes.com](http://www.nytimes.com)

Por más información sobre el tema periodismo de datos pueden consultar las páginas:

<http://datajournalismhandbook.org/1.0/en/>

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

[http://medialab-prado.es/article/consiguiendo\\_datos](http://medialab-prado.es/article/consiguiendo_datos)



## Capítulo IV

# Glosario de términos





## Glosario de términos

**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.



## Tabla de contenido

Introducción .....	7
1. ¿Qué son los Datos Abiertos? .....	7
2. ¿Qué datos de gobierno podrían ser abiertos? .....	8
Apertura de Datos .....	11
3. Clasificación de los datos abiertos .....	11
4. Procedimiento para la apertura de datos .....	14
Reutilización de datos .....	27
5. Ejemplos de reutilización de datos .....	27
6. Herramientas para el procesamiento de datos .....	33
7. Periodismo de datos .....	36
Glosario de términos .....	41
Tabla de contenido .....	44