

Directrices Técnicas para la Apertura de Datos Abiertos

1. Introducción

Este documento describe los aspectos técnicos básicos relativos a la publicación de datos abiertos, el cual contiene:

- Formatos inválidos para la publicación de Datos Abiertos.
- Criterios para la selección de formatos válidos
- Lista de formatos válidos más usuales, su aplicabilidad y las referencias a sus especificaciones.

El documento está orientado a personal técnico informático que incluya entre sus tareas la publicación de datos abiertos o la toma de decisiones sobre la publicación de datos abiertos de los organismos públicos del Uruguay.

En cada sección, se indicará si reviste carácter informativo u obligatorio.

Las **secciones informativas**, proveen contexto y definiciones para facilitar la comprensión del documento.

Las **secciones obligatorias**, prescriben directrices que deben ser cumplidas por los organismos públicos establecidos en el artículo 82 de la Ley N° 19.355, de 19 de diciembre de 2015.

2. Datos Abiertos.

La presente sección y todas sus subsecciones son informativas.

A los efectos de este documento, los **Datos Abiertos** son aquellos que pueden ser utilizados y distribuidos libremente por cualquier persona u organización, sin otra limitación que la atribución y la normativa nacional vigente.

2.1. Principios de la Publicación de Datos Abiertos

La publicación de Datos Abiertos se rige por ciertos principios orientados a la libertad de acceso y uso de los mismos. El presente documento se rige por los siguientes principios:

1. **Los datos deben ser completos:** Todos los datos públicos están disponibles. Los datos públicos son aquellos que no están sujetos a restricciones de privacidad, seguridad o privilegio.
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, sin ser modificados ni agrupados (exceptuando las modificaciones o exclusiones de datos, al amparo de la Ley N° 18.331, de 11 de agosto de 2008 de Protección de Datos Personales y la Ley N° 18.381, de 17 de octubre de 2008 de Acceso a la Información Pública, así como las disposiciones modificativas y concordante.
3. **Los datos deben ser oportunos:** Los datos están disponibles tan pronto como sea necesario para preservar su valor.
4. **Los datos deben ser accesibles:** Los datos están disponibles para cualquier usuario y propósito.
5. **Los datos deben ser procesables por máquinas:** Los datos están lo suficientemente estructurados como para permitir su procesamiento automático por dispositivos informáticos.
6. **El acceso a los datos no debe ser discriminatorio:** Los datos están disponibles para todos, sin requisitos de registro.
7. **Los datos se publican en formatos no-propietarios:** Los datos están disponibles en un formato sobre el que ninguna entidad tiene control exclusivo.
8. **Sin licencia restrictiva:** Los datos no están sujetos a ningún tipo de regulación de derechos, patentes o registros de marca. Se podrán permitir restricciones razonables de privacidad, seguridad o privilegio.

2.2. Publicación de Datos Abiertos.

Los Datos Abiertos se publican en **conjuntos de datos** o **datasets**. Estos conjuntos de datos están compuestos por 3 elementos básicos:

- Uno o más archivos en un **formato adecuado**.
- Uno o más archivos que describen el propio conjunto de datos y los datos contenidos en cada uno de los archivos. (**Metadatos**).
- Una **licencia** que indica cómo se debe realizar la atribución del conjunto de datos y si las hubiere, restricciones en su uso.

Los **formatos adecuados para la publicación de datos abiertos**, son aquellos que permiten realizar un procesamiento automático razonable de su contenido. Estos formatos tienen algunas características en común:

- Son formatos textuales (no binarios) en una codificación conocida, típicamente UTF-8.
- El contenido de los archivos (datos), está accesible sin necesidad de procesos complejos, costosos o imprecisos como por ejemplo, OCR o técnicas avanzadas de reconocimiento de lenguaje natural.

Los **metadatos** son datos que describen el conjunto de datos y su contenido. Estos incluyen información de las organizaciones y personas responsables de la publicación y generación de los datos, información estructural del contenido, e información que permite comprender y procesar ese contenido correctamente (semántica).

La **licencia**, es un documento de carácter jurídico que especifica bajo qué condiciones se puede hacer uso del conjunto de datos. La licencia se especifica a nivel de un conjunto de datos, de todo un organismo o incluso, de todo el gobierno. Las licencias que se usan para datos abiertos, tienden a asegurar que no hay restricciones de ningún tipo al uso de los datos y explican cómo se debe realizar la atribución.

2.3. Nivel de Apertura de un Conjunto de datos.

Los Datos Abiertos suelen ser clasificados según su nivel de apertura utilizando una categorización conocida como 5 estrellas:

* Los datos están publicados en la web en cualquier formato y con una licencia abierta asociada.

** Los datos están publicados en la web con una licencia abierta y en un formato estructurado que puede ser procesable por una computadora.

*** Los datos cumplen con el nivel de 2 estrellas, pero además el formato utilizado es no propietario y su acceso y procesamiento no dependen de ninguna organización pública o privada.

**** Los datos cumplen con el nivel de 3 estrellas, y además se utilizan estándares abiertos para identificar elementos abstractos o del mundo real y posibilitar que puedan ser referenciados mediante URIs.

***** Los datos cumplen con el nivel de 4 estrellas, y además están relacionados (enlazados) con datos de otros conjuntos de datos a través de sus URIs.¹

¹ Por más información al respecto, se puede consultar la siguiente página: <http://5stardata.info/es/> en donde se encontrarán ejemplos y ventajas y desventajas de cada nivel de la clasificación.

3. Publicación de Datos de Transparencia Activa como Datos Abiertos.

La presente sección y todas sus subsecciones son obligatorias.

3.1. Alcance

Los organismos alcanzados obligatoriamente en la presente sección son todos aquellos que están obligados por la N° 18.381, de 17 de octubre de 2008, de Acceso a la Información Pública y el Decreto N° 232/010, de 2 de agosto de 2010.

El artículo 82 de la ley N° 19.355 expresa que los organismos deberán:

“...como mínimo publicar en formato abierto, la información preceptuada por el artículo 5° de la Ley N° 18.381, de 17 de octubre de 2008, y por los artículos 38 y 40 del Decreto N° 232/010...”

Se entiende que este aspecto del artículo, refiere a la publicación como datos abiertos de los datos indicados.

Cada organismo, podrá decidir además, publicar como datos abiertos aquellos datos que considere útiles para algún sector particular de la sociedad aunque no estén comprendidos por el referido artículo.

3.2. Estructura y Localización de las Publicaciones.

Las publicaciones de datos abiertos deberán realizarse en al menos, uno de los siguientes sitios:

- En los sitios web institucionales de los organismos, que deberán contar con una sección propia para catalogar los datos abiertos siguiendo las recomendaciones de AGESIC establecidas para el Portal Tipo.
- El Catálogo Nacional de Datos Abiertos siguiendo las recomendaciones de AGESIC establecidas en la Guía para la publicación de datos abiertos de Gobierno, publicadas en el portal de AGESIC.

Para cada conjunto de datos publicado, deberá poder accederse mediante una o más URLs a sus diferentes elementos. Estos se encuentran descritos en la Sección 2.2 del presente documento:

- Los datos que conforman el contenido del conjunto de datos, bajo la forma de uno o más archivos.
- Los metadatos que describen ese contenido.
- La licencia de acceso contenida en la Sección 4 del presente documento.

Para que un conjunto de datos publicado sea considerado como datos abiertos, debe usar tanto para el contenido como para los metadatos, formatos adecuados para publicaciones que tengan como mínimo el nivel de apertura *** (3 estrellas).

3.3. Formatos de Archivos Adecuados y No Adecuados para Datos Abiertos.

Los formatos de publicación deberán seguir los criterios mencionados en la sección 2.2 del presente documento. Por este motivo, **los siguientes formatos son considerados inadecuados para la publicación de datos abiertos:**

- **Formatos gráficos con contenido textual.** Quedan excluidos de los formatos válidos para datos abiertos cualquier forma de documento textual escaneado, en cualquier formato gráfico y/o PDF.
- **Formatos de texto que no incluyan marcas o sólo incluyan marcas de presentación.** Quedan excluidos de los formatos válidos para datos abiertos los formatos de texto plano puro (texto en Ascii). Por marcas de presentación se entiende cualquier indicación de estructura de documento (título, párrafo, etc.) o formateo (itálica, negrita, etc.). De esta forma, quedan excluidos los formatos de tipo markdown o similares o incluso, html sin RDF-a incluido o xml sin xml-schema o con xml-schema que incluya sólo especificaciones de formato.
- **Formatos de archivo que necesiten de programas o bibliotecas con licenciamiento propietario para ser accedidos o procesados.** De esta forma quedan excluidos de los formatos válidos para datos abiertos los formatos generados por aplicaciones específicas como .doc, .xls, etc.

Existen dos tipos de formatos adecuados para la publicación de datos abiertos:

- **Formatos Generales.** Son formatos que pueden especializarse para usos más específicos. Los dos más habituales son XML², y RDF³.
- **Formatos Específicos.** Son formatos basados en los formatos generales, pero que ya están especializados para un uso determinado.

En la siguiente tabla se puede ver una lista de formatos usuales para la publicación de datos abiertos. Estos **formatos son considerados adecuados** de acuerdo al presente documento:

Nombre	Uso esperado	Referencias.
CSV	Datos con estructura relacional (Tablas)	RFC4180
XML	Datos con estructura jerárquica (Arboles)	XML 1.0, XML 1.1
JSON	Datos con estructura jerárquica (Arboles)	
RDF	Datos con cualquier estructura. (Grafos)	RDF 1.1

² [Bray et al. 2006, Bray et al. 2008]

³ [Wood, Lanthaler, y Cyganiak]

4. Licencia de Datos Abiertos de Uruguay

Todo conjunto de datos abiertos que sea publicado por un organismo del Estado deberá hacer referencia en la declaración de sus metadatos al siguiente texto de licencia.

Licencia de Datos Abiertos - Uruguay

Versión 0.1

Autorización. Los datos abiertos y los metadatos abiertos proporcionados, para uso comercial y no comercial podrán ser:

- enajenados, reproducidos, distribuidos, publicados, traducidos, adaptados, transformados, comunicados o puestos a disposición del público mediante cualquier forma o procedimiento;
- fusionados con datos propios y de terceros, pudiendo combinarse para formar conjuntos de datos nuevos e independientes;
- integrados en los procesos de negocio, productos y aplicaciones internas y externas en las redes electrónicas públicas y privadas.

Nota de Origen. A los efectos de utilizar los datos abiertos y metadatos abiertos en los términos referidos en el artículo anterior se deberá en todos los casos citar la siguiente información:

- el nombre del proveedor,
- la referencia a la "Licencia de Datos Abiertos – Uruguay".
- la referencia al conjunto de datos que está siendo utilizado.

Los cambios, las ediciones, los nuevos diseños o modificaciones deben ser mencionados en la nota de origen.

Bibliografía

[Tauberer] Tauberer, Joshua. *Open Government Data: The Book*. Second Edition. Lulu, 214d. C. <https://opengovdata.io/>.

[Dietrich et al.] Dietrich, Daniel, Jonathan Gray, Tim McNamara, Antti Poikola, Rufus Pollock, Julian Tait, y Ton Zijlstra. «Open Data Handbook.» Accedido 18 de mayo de 2016. <http://opendatahandbook.org/guide/>.

[Bray et al. 2006] Bray, Tim, Jean Paoli, C.M. Sperberg-McQueen, Eve Maler, François Yergeau, y John Cowan. «Extensible Markup Language (XML) 1.1 (Second Edition).» W3C Recommendation. W3C - World Wide Web Consortium, septiembre de 2006. <http://www.w3.org/TR/2006/REC-xml11-20060816/>.

[Bray et al. 2008] Bray, Tim, Jean Paoli, C. M. Sperberg-McQueen, Eve Maler, y François Yergeau. «Extensible Markup Language (XML) 1.0 (Third Edition), W3C Recommendation.» W3C, 2004. <http://www.w3.org/TR/2004/REC-xml-20040204>.

[Wood, Lanthaler, y Cyganiak] Wood, David, Markus Lanthaler, y Richard Cyganiak. «RDF 1.1 Concepts and Abstract Syntax.» W3C Recommendation. W3C, febrero de 2014. <http://www.w3.org/TR/rdf11-concepts/>.