



Recomendaciones técnicas para estandarizar la publicación de Datos Abiertos

Introducción

Este documento está dirigido a equipos técnicos encargados de la gestión de los datos para su apertura y publicación.

La normativa vigente ([artículo 82 de la Ley 19.355](#)) establece que las Instituciones del Estado deben publicar los datos públicos en formatos abiertos. Otras normas que establecen obligaciones vinculadas a la recopilación y publicación de datos son, el [Art. 73 de la Ley 20.212](#) que obliga a entidades públicas estatales y no estatales a publicar datos de cartografía y el artículo 11 de la [Ley 19. 846](#) de 2019 que obliga a los organismos públicos a recopilar información sobre sexo, género, ascendencia étnico-racial, etc. Los equipos técnicos responsables de la gestión de los datos y de la publicación de los datos abiertos deben adaptar sus procesos para garantizar el cumplimiento de dicha normativa.

Es importante destacar que el valor de estos procesos se encuentra en el uso de los datos dentro y fuera del gobierno, ya sea para la toma de decisiones, para la mejora en la entrega de servicios, para impulsar procesos de innovación y desarrollo de nuevas soluciones, para facilitar el acceso a la información pública o para promover procesos de participación ciudadana entre otros.

La adopción de estándares de publicación facilita la interoperabilidad, la integración e interpretación de los datos, facilitando así el acceso y el uso de los mismos.





El presente documento contiene un conjunto de Recomendaciones Técnicas para la identificación de los archivos de fuentes de datos, para la estructuración de archivos tabulares y estándares para diferentes tipos de datos.

Definiciones

A efectos de facilitar el entendimiento del presente documento se establecen a continuación un conjunto de definiciones.

Conjunto de datos

De acuerdo con la [norma técnica](#) aprobada por el [Decreto reglamentario 054/017](#) del febrero 2017 se define al conjunto de datos como una agrupación de uno o más archivos de datos, metadatos y una licencia abierta que habilita el uso de datos de acuerdo con ciertas condiciones.

“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.”
-

Recursos

Se define como un recurso de datos (o recurso) a cualquier fuente de datos que se encuentre publicada en el Catálogo Nacional de Datos Abiertos, sea accesible





mediante un enlace (URL Uniform Resource Locator) y forme parte de una entidad mayor definida como conjunto de datos (El término en inglés, dataset; también es usado habitualmente como sinónimo de conjunto de datos). El catálogo acepta otros tipos de recursos que complementen la información como pueden ser:

Metadatos

Diccionario de variables

Fichas metodológicas

Otras Licencias abiertas no disponibles en el catálogo

Distribuciones

Un recurso de datos puede estar disponible en múltiples distribuciones dependiendo del formato en que se publique, esto significa que una misma fuente de datos se puede distribuir en diferentes formatos como por ejemplo CSV, XML, JSON, KML, a través de servicios web, etc.

Datos

De acuerdo con la Real Academia Española desde el punto de vista informático un dato es:

“Información dispuesta de manera adecuada para su tratamiento por una computadora.”

Los datos son esenciales para el conocimiento y la toma de decisiones. Se utilizan en todos los ámbitos de la vida, desde la ciencia y la tecnología hasta la economía y la política. Los datos nos permiten comprender el mundo que nos rodea, identificar problemas y encontrar soluciones.





Recomendaciones para la publicación de datos abiertos

1. Fuentes de datos

a. Recomendaciones generales para archivos de datos

En caso de que la fuente de datos sea un archivo este deberá cumplir con las siguientes condiciones:

- Nombre del archivo
 - El nombre del archivo solo acepta números, letras y guion del medio, excluyendo caracteres extraños como tildes, diéresis, virgulilla, etc.
 - Si el nombre tiene más de una palabra estas deberán separarse con guión medio.
 - El guión no debe utilizarse como primer o último carácter en el nombre del archivo.
 - El dominio del contenido del archivo debe reflejarse en el nombre del archivo.
 - El versionado del archivo debe ubicarse al final, separado del nombre por un guión medio.

Ejemplos correctos: 1) rendicion-cuentas-2021-v001.xml, 2) rendicion-cuentas-2021-v001.csv

- Tamaño del archivo
 - Se recomienda que los archivos de datos no superen los 2Gb de datos.
 - Si el tamaño de los archivos de datos supera los 2Gb de datos, se debe considerar la fragmentación del archivo en la





cantidad de archivos necesarios que cumplan la condición anterior.

- Si los archivos son fragmentados se recomienda nombrarlos en el orden que corresponda, agregando un número ordinal de dos dígitos al final del nombre. Ejemplo: rendicion-cuentas-2021-v001-01.csv

- Archivos comprimidos

- La publicación de archivos de datos comprimidos sólo se recomienda para facilitar la descarga de grandes volúmenes de datos, siempre que se cumpla la existencia de alternativas para la descarga parcial de los datos.

- Recursos

- Los recursos incluidos en un conjunto de datos no pueden incluir valores que sean el resultado de operaciones realizadas a partir del procesamiento de los datos incluidos en el archivo. Ejemplo: totales u otro valor que refiera a un nivel de granularidad diferente.

b. Recomendaciones específicas para archivos de datos tabulares

- La publicación de datos abiertos debe hacerse utilizando la codificación UTF-8.
- La primera fila debe contener los nombres de las columnas.
- Las columnas pueden crecer a lo largo del ciclo de vida del conjunto de datos hacia la derecha y al final de la última columna existente; para mantener compatibilidad con procesos anteriores.





- Los nombres de las columnas deben respetar alguno de los siguientes formatos: Mayúsculas/Minúsculas (Camel Case).

Ejemplo: 1) Lower Camel Case = fechaFinal o 2) Upper Camel Case = FechaFinal

- Texto en minúsculas separando palabras con guión bajo (Snake Case). Ejemplo: fecha_final
- Los nombres de las columnas no deben incluir espacios ni caracteres especiales a excepción del guión bajo.
- Los datos deben crecer en forma vertical, agregando nuevos datos en la primera fila vacía después de la última fila con datos.
- Se recomienda que la cantidad de filas de datos no supere el millón.
- Si la cantidad de filas supera el millón, se debe fragmentar el archivo en la cantidad de archivos necesarios que cumplan la condición anterior.
- Se recomienda que los datos tabulares se publiquen en formato de valores separados por coma (CSV) según lo especificado por la RFC 4180.
- En los casos que existan campos que contengan valores múltiples, los valores deberán separarse con un carácter “|”

Ejemplo: 2023-01-01,234|405|341|2|89,”Juan”,”Pérez”,

c. Servicios web

- Si la fuente de datos se hace disponible mediante un servicio web o API, la URL debe cumplir con [la RFC1738](#).

protocolo://<servidor>:<puerto>/<ruta>?<parámetros>





Ejemplo de URL:

<https://www.comprasestatales.gub.uy/ws/ActualizacionCatalogoWS/ActualizacionCatalogoService.ActualizacionCatalogoServiceHttpSoap11Endpoint?wsdl>

2. Estándares para tipos de datos

a. Datos Temporales

Los datos de tipo fecha y/o tiempo, se deben publicar siguiendo el estándar ISO 8601.

2020-07-10 15:00:00 -- AAAA-MM-DDTHH:MM:SS





Dónde:

AAAA: representa el año con 4 dígitos. Ej. 2022

MM: representa el mes con 2 dígitos: Ej. 01 para representar a enero

DD: representa los días con 2 dígitos: Ej. 31.

Con respecto al tiempo se utiliza el mismo estándar en formato extendido:

Thh:mm:ss la letra T se utiliza como separador y puede ser omitida.

HH: la hora representada con 2 dígitos de 00 a 23.

MM: son los minutos de 00 a 59.

SS: son los segundos de 00 a 59.

Para la representación de períodos de fecha se utilizará el carácter de “:” para separar la fecha de inicio de la fecha de finalización.

2020-07-10:2021-07-09 para indicar el periodo que inicia el 10 de julio de 2020 y finaliza el 9 de julio de 2021.

b. Datos Numéricos

- No utilizar separador de miles.
- Utilizar el carácter de punto para separar decimales.
- Limitar la cantidad de decimales a la mínima expresión necesaria.
- Se recomienda utilizar dos decimales si la precisión del dato lo permite.
- El signo de menos se utilizará para determinar valores negativos.

c. Datos de tipo Moneda





- No utilizar separador de miles.
- Utilizar el carácter de punto para separar decimales.
- Utilizar dos decimales.
- El signo de menos se utilizará para determinar valores negativos.
- No utilizar identificador para el tipo de moneda.
- Para representar más de un tipo de moneda se debe incluir un campo que especifique el tipo de moneda.

d. **Datos de Edad o rango etario**

- La publicación de datos sobre edad se puede expresar con un número entero de hasta 3 dígitos. Los rangos etarios se deben generar teniendo en cuenta la no identificación de las personas.
- Si se publican fechas de nacimiento se respetará el formato para el tipo de datos fecha, definido anteriormente y considerando que no se incumpla con la normativa vigente para la protección de datos personales y que puedan hacer identificable a una persona.

3. **Datos referenciales**

Los datos referenciales son aquellos que definen los rangos de valores permitidos para un tipo de dato. Cuando los datos se publican utilizando datos referenciales estandarizados habilitan la interoperabilidad y facilitan su reutilización.

4. **Socio - demográficos**

a. **Territorio**





- **Países:** Codificación de países de acuerdo con la norma ISO 3166-1:2007. Descargar tabla de codificación desde el Catálogo Nacional de Datos Abiertos, conjunto: [Datos referenciales](#).
- **Departamentos:** Codificación dada en el modelo de direcciones geográficas del Uruguay a los departamentos de la República Oriental del Uruguay, de acuerdo con la norma ISO 3166-2:2007. Descargar tabla de codificación desde el Catálogo Nacional de Datos Abiertos, conjunto: [Datos referenciales](#).
- **Localidades:** Listado de localidades de la República Oriental del Uruguay. Las coordenadas de las localidades se proyectan en el sistema EPSG 32721. Descargar tabla de codificación desde el Catálogo Nacional de Datos Abiertos, conjunto: [Localidades del Uruguay](#).
- **Códigos postales:** El código postal es un identificador numérico o alfanumérico compuesto de una serie de letras y/o cifras e identifica un punto de entrega o un conjunto de puntos de entrega. Los datos para considerar cuando se use esta clasificación son los brindados por El Correo. Descargar tabla de codificación desde el Catálogo Nacional de Datos Abiertos, conjunto: [Código Postal](#).

b. **Sexo y/o Perspectiva de género**

- **Sexo:** Codificación del sexo entendido como “el conjunto de características biológicas que definen al espectro de humanos” (Organización Mundial de la Salud). Para fines administrativos o legales se entiende por sexo a aquel con el que nació la persona. Descargar tabla de codificación desde el Catálogo Nacional de Datos Abiertos, conjunto: [Datos referenciales](#).





- **Perspectiva de género:** Codificación de la perspectiva de género entendiendo como la construcción cultural que alude a diferentes rasgos, conductas, actividades y funciones atribuidas a las personas, con independencia del sexo biológico. Descargar tabla de codificación desde el Catálogo Nacional de Datos Abiertos, conjunto: [Datos referenciales.](#)





c. **Ascendencia racial y/o étnica**

- **Ascendencia racial:** Codificación de la ascendencia racial realizada por el Instituto Nacional de Estadísticas. Descargar tabla de codificación desde el Catálogo Nacional de Datos Abiertos, conjunto: [Datos referenciales.](#)

5. Vías de circulación

a. **Direcciones geográficas**

- La base de direcciones fue generada inicialmente por el Correo Uruguayo en base a información propia y de otras instituciones. Contiene direcciones oficiales (generadas por los gobiernos departamentales) y no oficiales (existentes en la realidad). Se implementó basándose en el Modelo de Direcciones Geográficas del Uruguay.
- Descargar tabla de codificación desde el Catálogo Nacional de Datos Abiertos, conjunto: [Direcciones Geográficas del Uruguay.](#)

6. Estructura del Estado

a. **OID**

Un OID es un identificador de objetos y se define de acuerdo con una asignación jerárquica, establecida por la Organización Internacional de Normalización (ISO), basada en la Norma ISO/IEC 8824-1.

La UNAOID (Unidad Nacional de Asignación de OID), a través de estos identificadores busca proporcionar una base estandarizada para la interoperabilidad de los sistemas de información en la Administración Pública, así como en el resto de los sectores de la sociedad.





La raíz nacional asignada por ISO e ITU-T a la UNAOID, de acuerdo con los procedimientos ISO/IEC JTC 1/SC 6/WG 9 y ITU-T Q.12/17 respectivamente, corresponde a: 2.16.858.

Descargar tabla de codificación desde el Catálogo Nacional de Datos Abiertos, conjunto: [UNAOID - Listado de OIDs de la rama organización](#).

b. **Organismos del presupuesto Nacional**

Para la clasificación de organismos incluidos en el presupuesto nacional se utiliza la codificación del Sistema Integrado de Información Financiera (SIIF).

Esta codificación se encuentra publicada por la Oficina de Planeamiento y Presupuesto (OPP) en el Catálogo Nacional de Datos Abiertos.

Descargar tabla de codificación desde el Catálogo Nacional de Datos Abiertos, conjunto: [Organismos del Presupuesto Nacional](#).

