



Uruguay
Presidencia



Glosario general de Datos Abiertos

Área: Sociedad de la Información

Versión: 1.0

Año: 2024-01



Contenido

1. Terminología	3
2.1. Dato	3
2.2. Dato Público.....	3
2.3. Dato Abierto	3
2.4. Conjunto de datos	4
2.5. Licencia.....	4
2.6. Recurso web	5
2.7. Metadato	5
2.8. Formato de archivo	5
2.9. Formato abierto.....	6
2.10. Distribución.....	6
2.11. Anonimización	6
2.12. Catálogo Nacional de Datos Abiertos	7
2.13. Categoría.....	7
2.14. Visualización	7
2.15. Aplicación	7
2.16. Usabilidad.....	8
2.17. CKAN	8
2.18. Harvest.....	8
2.19. API	8



1. Terminología

El siguiente documento presenta un diccionario con todos los términos utilizados en los diferentes procedimientos y guías de trabajo con Datos Abiertos.

2.1. Dato

Según la [Real Academia Española](#) “un dato es la información sobre algo concreto que permite su conocimiento exacto o sirve para deducir las consecuencias derivadas de un hecho”. En otras palabras, un dato es la representación simbólica de un atributo o variable cuantitativa o cualitativa. Los datos aislados no contienen información relevante, es a partir del análisis de los datos y del contexto que se genera información.

2.2. Dato Público

Son datos que no están sujetos a restricciones de privacidad, seguridad o privilegio y que se encuentran clasificados como información pública por la [Ley de Acceso a la Información Pública](#). No existen restricciones en cuanto al formato en el cual es distribuido.

2.3. Dato Abierto

Según la [Open Data Handbook](#) “Los datos abiertos son datos que pueden ser utilizados, reutilizados y redistribuidos libremente por cualquier persona, y que se encuentran sujetos, cuando más, al requerimiento de atribución y de compartirse de la misma manera en que aparecen.” Entonces los datos abiertos además de ser datos públicos deben estar publicados bajo una licencia abierta y cumplir con los 8 principios de datos abiertos definidos por la [Open Government Data](#) los cuales son; completos, primarios, oportunos, accesibles, procesables por máquinas, no discriminatorios y no propietarios.



2.4. Conjunto de datos

Un conjunto de datos es una colección o estructura de datos relacionados bajo un mismo criterio o tema que ayuda a gestionarlos, habitualmente se encuentran tabulados.

A nivel del [Catálogo Nacional de Datos Abiertos](#) un conjunto de datos está formado por la información la cual describe al conjunto de datos (metadata) y por sus recursos; datos y metadatos (datos descriptivos o diccionario de datos). Cada recurso de un conjunto puede ser publicado en uno o más formatos (distribuciones) diferentes.

2.5. Licencia

Es un documento de carácter jurídico que especifica bajo qué condiciones se puede hacer uso de los conjuntos de datos. Los organismos públicos en Uruguay deben utilizar la [Licencia de Datos Abiertos de Uruguay](#), no obstante, existen otras licencias que pueden ser utilizadas para licenciar datos abiertos como:

- **Creative Commons 3.0:** Posibilita la distribución, modificación, creación y reuso con fines comerciales, siempre y cuando se dé crédito a la fuente de creación original.
- **GNU GPL:** Garantiza a los usuarios finales, estos sean personas, organizaciones, compañías, la libertad de utilización, copia y modificación de la obra.
- **Open Database License (ODbL):** Orientada al uso de bases de datos, permite la copia, distribución, modificación y creación y reuso con fines comerciales, siempre y cuando se dé crédito a la fuente de creación original y se sostenga la misma licencia para cualquier creación derivada.



2.6. Recurso web

Un recurso web es cualquier elemento que se encuentre en línea y que se pueda ver, leer, escuchar o interactuar, para un catálogo de datos los recursos son los contenedores de archivos que permiten a los usuarios acceder a los datos publicados.

2.7. Metadato

El término refiere a datos que describen datos. En el caso del [Catálogo Nacional de Datos Abiertos](#) el término hace referencia a distintas descripciones, en primer lugar, a los atributos que describen un conjunto de datos (metadata), también hay metadatos que describen a los recursos de un conjunto y por último existen los metadatos que describen el contenido de los archivos de datos abiertos. Todos estos metadatos ayudan a comprender mejor la publicación y la correcta interpretación de los datos en su procesamiento o reuso.

2.8. Formato de archivo

Un formato de archivo especifica cómo se define y almacenan los datos para una aplicación determinada. Cada formato tiene características diferentes que le permiten almacenar determinados tipos de datos conforme las necesidades de quien crea y/o gestiona esos archivos. A continuación, algunos ejemplos:

- xlsx (Planilla electrónica del nuevo Microsoft Excel)
- csv (Archivo de texto con un formato de valores separados por comas)
- odx (planilla electrónica de Libre Office)
- docx (Documento de texto del nuevo Microsoft Word)
- pdf (formato de documentos creado por Adobe)



- xml (archivo de lenguaje marcado extensible y estructurado)
- Json (formato utilizado como alternativa al XML para la transferencia de datos estructurados)

2.9. Formato abierto

Cuando hablamos de formato abierto hacemos referencia a la especificación del tipo de archivo el cual debe estar disponible públicamente y de manera gratuita para su uso y no impone ninguna restricción según la [Real Academia Española](#).

El concepto de “abierto” está definido por la [Open Knowledge Foundation](#), donde dice que un formato abierto es aquel que no impone restricciones, monetarias o de otro tipo sobre su uso y puede procesarse completamente con al menos una herramienta de software libre y de código abierto.

2.10. Distribución

En el contexto de los Datos Abiertos este término se refiere a los diferentes formatos que están publicados los archivos de datos. Cada distribución permite la descarga de los datos en distintos formatos.

Ejemplo: Puede publicar un conjunto de datos llamado Horarios de transporte, cuyos datos se publiquen en diferentes formatos: CSV, XLSX, XML y JSON. Su conjunto de horarios de transporte contendrá un recurso llamado horarios de transporte de verano y un recurso llamado horarios de transporte de invierno y cada recurso estará distribuido en cuatro formatos.

2.11. Anonimización

La anonimización es una técnica de tratamiento de datos que elimina o modifica los datos personales identificables para obtener datos anónimos que no se pueden



asociar con ninguna persona de forma directa o indirecta.

2.12. Catálogo Nacional de Datos Abiertos

El [Catálogo Nacional de Datos Abiertos](#) es una herramienta que permite acceder a datos abiertos de organismos públicos, academia, organizaciones de sociedad civil y empresas privadas. Cualquier persona puede utilizar los datos publicados libremente para contar historias, desarrollar investigaciones, visualizaciones, aplicaciones cívicas y emprendimientos.

Se accede desde <https://catalogodatos.gub.uy>

2.13. Categoría

La [Real Academia española](#) define categoría como: “Condición social de unas personas respecto de las demás. Calidad o importancia de alguien o algo”. En el Catálogo Nacional de Datos abiertos una categoría es una clasificación utilizada para los conjuntos de datos, que permite agruparlos por distintas temáticas. Cada categoría es una temática distinta.

2.14. Visualización

Una visualización es una representación gráfica de los datos publicados. Se utilizan elementos visuales como gráficos, tablas o mapas y sirven como vista previa de los datos que cuenta un conjunto.

2.15. Aplicación

El término de “Aplicación” en el [Catálogo Nacional de Datos Abiertos](#) es utilizado para asociar contenidos a algún conjunto de datos abiertos publicado en el catálogo, pueden ser aplicaciones para dispositivos móviles, visualizaciones de datos, sitios web, informes, investigaciones, etc.



2.16. Usabilidad

El organismo de estandarización ISO (International Standardisation Organization) proporciona la siguiente definición para Usabilidad “La medida en la cual un producto puede ser usado por usuarios específicos para conseguir objetivos específicos con efectividad, eficiencia y satisfacción en un contexto de uso especificado.”

2.17. CKAN

Es la herramienta en la cual está implementado el [Catálogo Nacional de Datos Abiertos](#). Es un desarrollo de código abierto desarrollado por la [Open Knowledge Foundation](#) y mantenido por [CKAN.org](#).

2.18. Harvest

Harvest es una palabra en ingles que significa cosecha. La misma en el [Catálogo Nacional de Datos Abiertos](#) hace referencia a una funcionalidad del CKAN que facilita la recolección automática de datos y metadatos desde otros catálogos externos, añadiendo distintos conjuntos de datos a su propio Catalogo. Por eso se le llama cosecha, porque permite cosechar conjuntos generados (publicados) y mantenidos en fuentes externas.

2.19. API

API sigla significa que procede de la lengua inglesa y hace alusión a (Application Programming Interface). Es una interfaz de comunicación entre componentes de software, utilizado por sistemas informáticos diferentes para comunicarse entre ellos. [CKAN cuenta con una API propia](#) que facilita la publicación y el consumo automatizado de datos.