

GUÍA PARA EL USO Y LA PUBLICACIÓN DE METADATOS

INTRODUCCIÓN

VERSIÓN

Esta guía es compatible con la **versión 1.1 del Perfil Nacional de Metadatos para Datos Abiertos** de la Administración Pública Nacional.

OBJETIVO

El objetivo de esta guía es que los organismos de la Administración Pública Nacional alcanzados por el Decreto N° 117/2016 del 12 de enero de 2016, encuentren todos los recursos necesarios para **crear su catálogo de datos abiertos** según el Perfil Nacional de Metadatos para Datos Abiertos impulsado por la Dirección Nacional de Datos e Información Pública.

El estándar desarrollado y las recomendaciones vertidas en este documento se basan en estándares usados a nivel nacional e internacional, y en la experiencia de trabajo del equipo de Datos de la Secretaría de Gobierno de Modernización de la Nación.

Esta es **una guía colaborativa y en progreso**. Valoramos, y alentamos, a organizaciones y ciudadanos a plantear ideas, sugerencias, y comentarios que nos ayuden a crear un mejor documento.

Este documento se complementa con la **Guía para la publicación de datos en formatos abiertos** y la **Guía para la identificación y uso de entidades interoperables**.

¿QUÉ SON LOS METADATOS?

Los metadatos son los datos sobre los datos. Es decir: **elementos descriptivos que dan contexto** a un conjunto de datos, y acercan al usuario la información necesaria para entenderlos y usarlos eficazmente.

Un **título** y una breve **descripción** son los metadatos básicos que cualquier conjunto de datos a publicar debería tener.

Luego, existen muchos otros elementos que ayudan al usuario a utilizar los datos correctamente. Por ejemplo:

- **Nombre, tipo de datos y descripción de los campos:** ¿qué significa cada campo? ¿qué datos puedo encontrar en esa columna? ¿qué dicen y qué *no* dicen esos datos, cómo debo leerlos?
- **Palabras clave:** clasifican a un dataset como perteneciente a un conjunto de tópicos.
- **Tema:** clasifican a un dataset como perteneciente a un determinado tema, dentro de una jerarquía temática.
- **Fecha de publicación:** ¿cuándo se publicó por primera vez este dataset?
- **Fecha de última modificación:** ¿cuándo se actualizó por última vez este dataset?
- **Frecuencia de actualización:** ¿cada cuánto se actualiza este dataset?
- **URL de descarga:** ¿cómo dispongo de los datos, desde dónde puedo descargarlos?

Una lista de metadatos, junto con las instrucciones de cómo deben utilizarse, es un **perfil de metadatos**.

¿CÓMO SE PUBLICAN LOS METADATOS?

Los metadatos se publican en distintos formatos, estructuras y modalidades dependiendo del objetivo, la temática que abordan o las aplicaciones automáticas que los consuman.

Si se desea documentar activos de datos, **una publicación elemental sería un documento de texto** (ver **Anexo VI - Ejemplo de metadatos como texto**) con una lista de datasets que ofrezca el título y la descripción de cada uno de ellos, y de los recursos que lo componen.

Sin embargo, **las computadoras no pueden leer fácilmente documentos de texto**. No se podrían utilizar aplicaciones automáticas para validar, traducir, re-indexar metadatos en otros catálogos y buscar datasets en ellos.

Un catálogo de datos abiertos se crea para organizar sistemáticamente colecciones de datasets y permitir el desarrollo de todo tipo de aplicaciones sobre ellos. Esto requiere el uso de otros formatos más estructurados.

La potencial reutilización de los conjuntos de datos y la posibilidad de que sean correctamente federados desde el Portal Nacional de Datos Abiertos dependerá de la calidad de sus metadatos y de su formato. Adoptar y/o desarrollar estándares y vocabularios controlados es importante para la lectura e interpretación de los conjuntos de datos por personas y por aplicaciones informáticas.

Para esto, **los catálogos de datos publican sus metadatos en un formato estructurado (JSON)** respetando un determinado perfil estandarizado. Recomendamos ver un ejemplo en JSON de los metadatos de un catálogo de datos en el **Anexo IV - Ejemplo de data.json**.

En el resto de este documento detallamos las características de los estándares y vocabularios controlados adoptados para catálogos de datos, datasets y distribuciones.

PÚBLICO OBJETIVO DE ESTA GUÍA

Esta guía es para:

- **Organismos de la Administración Pública Nacional** que quieren implementar un catálogo de datos abiertos
- Quienes implementan un Portal Andino, otro portal web, un catálogo en Excel o en JSON.
- **Otros organismos del Poder Legislativo o Judicial** que sin estar alcanzados por el Decreto 117/2016, adhieren a la Política de Apertura de Datos propuesta por el Poder Ejecutivo.
- **Gobiernos subnacionales** (provincias y municipios) que desean publicar sus catálogos de datos abiertos siguiendo los mismos estándares que el Gobierno Nacional y otros gobiernos subnacionales.
- **Universidades, ONGs, cámaras empresarias y otros actores de la sociedad civil** que adhieren a la Política de Datos Abiertos y desean publicar datos siguiendo los mismos estándares que el sector público, con la visión de construir una interoperabilidad creciente entre datos del gobierno nacional y de otros actores de la sociedad civil.

CREAR UN CATÁLOGO DE DATOS ABIERTOS

Para cumplir con los estándares de documentación de datos abiertos, **los organismos necesitan publicar un catálogo en formato XLSX o JSON en una URL pública dentro de su propiedad digital**, y comunicarla al equipo de Datos de la Secretaría de Gobierno de Modernización para su indexación en datos.gob.ar.

Si el catálogo cumple con el perfil de metadatos, se le asignará un identificador (Ej.: `agroindustria`, `energia`, `sspre`) y será federado en datos.gob.ar bajo el nombre del organismo responsable de su publicación.

Los datasets federados desde ese catálogo, serán accesibles públicamente desde `http://datos.gob.ar/dataset?organization={identificador}`.

EXCEL

El catálogo se puede crear en Excel, a partir de este modelo diseñado para cumplir con el perfil de metadatos. El modelo contiene una serie de validaciones para evitar errores en su uso, pero debe tenerse en cuenta que:

- En las hojas de distribución (`distribution`) y campo (`field`) **sólo deben asignarse los ids de los dataset o distribuciones "padre" que los incluyen**, el título se trae automáticamente de la hoja correspondiente con una fórmula.
- **Todas las columnas sean de tipo "texto"** (no "número") salvo aquellas que en el modelo contienen fórmulas para asegurar que una distribución o campo documentado, sean asignados al dataset o distribución correctos.

- Especialmente las columnas de identificadores (Excel a veces las convierte a número cuando el identificador elegido no contiene letras u otros caracteres).
- **La hoja "referencias" explica el código de colores** para saber qué campos son obligatorios, recomendados, opcionales o avanzados / especiales.
- El modelo contiene comentarios en las celdas para explicar brevemente qué se debe poner en cada una, pero en esta guía se puede consultar la referencia detallada.

PORTAL ANDINO

El **Portal Andino** es una herramienta especialmente diseñada para facilitar la publicación y apertura de datos y **cumple con el perfil de metadatos de la Administración Pública Nacional** propuesto en esta guía.

Incluye formularios web para la carga de datos y metadatos, y produce automáticamente el catálogo en versión XLSX y JSON.

El Portal debe publicarse en un dominio de la forma *http://datos.[entidad].gob.ar*. Esto asegura la publicación automática de los metadatos en formato *data.json* en *http://datos.[entidad].gob.ar/data.json* y XLSX en *http://datos.[entidad].gob.ar/catalog.xlsx*.

OTROS CATÁLOGOS

Sin embargo, los organismos de la APN **pueden optar por desarrollar sus propias implementaciones del Perfil de Metadatos**.

Quienes opten por una solución alternativa al Excel o al Portal Andino para publicar sus datos, deberán publicar los metadatos de su catálogo en un archivo *data.json* en una URL como la siguiente: *http://datos.[entidad].gob.ar/data.json*.

En [entidad] recomendamos usar un nombre simple y breve que represente a la organización que publica el catálogo (por ejemplo: *datos.jus.gob.ar*, *datos.tucuman.gob.ar*, *datos.pilar.gob.ar*).

Este archivo deberá estar construido tal como se puede ver en el ejemplo del **Anexo IV - Ejemplo de data.json**, respetando el Perfil de Metadatos de la Administración Pública Nacional tal como se lo describe más adelante en la sección "*Campos del perfil*".

PERFIL DE METADATOS

ESTÁNDAR USADO

El **Perfil Nacional de Metadatos para Datos Abiertos es una extensión del estándar DCAT - AP**, usado por los países de la Unión Europea. DCAT es un vocabulario controlado definido por la W3C, ampliamente usado a nivel global para la descripción de catálogos de datos.

Según la W3C: "Mediante la utilización de DCAT para describir datasets en catálogos de datos, quienes publican aumentan la posibilidad de descubrimiento (*discoverability*) y permiten a aplicaciones informáticas consumir metadatos de manera simple desde múltiples catálogos. Además permite la publicación descentralizada de catálogos y favorece la búsqueda *federada* de datasets a través de varios sitios."

El perfil de metadatos propuesto para la Administración Pública Nacional se compone de 3 clases principales (*Catalog*, *Dataset* y *Distribution*) y 2 auxiliares (*Field* y *Theme*) que se relacionan según el siguiente esquema:

□

A continuación, describimos los metadatos que el *data.json* debe contener, para cada una de estas clases.

CAMPOS DEL PERFIL

Catálogo (*catalog*)

El *data.json* de quienes usen el portal Andino, se generará a través de los formularios completados, publicándose automáticamente en *http://datos.[entidad].gob.ar/data.json*.

Ejemplos de metadatos de un **catálogo**:

- Catálogo en texto

- Catálogo en JSON

Metadatos necesarios para describir el catálogo, que un *data.json* debe contener:

Nombre	Requerido	Descripción	Ejemplo	Variable (data.json)	Tipo (data.json)
Identificador	Recomendado	En Argentina, es el identificador único del catálogo dentro de la Red de Nodos de Datos Abiertos de la Administración Pública Nacional. Este identificador es otorgado por la Dirección Nacional de Datos e Información Pública cuando un nuevo nodo pide ser incorporado a la red para su federación en el nodo concentrador de datos abiertos de la APN (http://www.datos.gob.ar). El identificador debe ser una o más palabras en minúsculas, separadas con guiones medios, sin usar caracteres especiales. Identifica en forma breve, sucinta y declarativa al nodo.	"enacom" "energia" "desarrollo-social" "justicia" "arsat"	identifier	String
Nombre	Sí	Nombre dado al catálogo. Debe ser claro, breve y lo suficientemente abstracto como para abarcar la multiplicidad de datasets que contiene.	Datos Argentina	title	String
Descripción	Sí	Descripción del contenido del catálogo.	Portal de Datos Abiertos del Gobierno de la República Argentina	description	String
Autor	Sí	Responsable de la publicación del catálogo.	Ministerio de Modernización	publisher -> name	String
Correo electrónico del autor	Sí	Correo electrónico de contacto del responsable de la publicación del catálogo.	datos@modernizacion.gob.ar	publisher -> mbox	String
Datasets	Sí	Contiene una lista de los datasets que forman parte del catálogo.	[{...}, {...}]	dataset	Array
Fecha de creación o publicación	Recomendado	Fecha de creación o publicación del catálogo. Se escribe según el formato ISO-8601, tipeado como fecha simple o fecha con hora, con el mayor detalle posible que sea relevante para el dataset.	"2016-04-14T19:48:05.433640" para especificar fecha y hora "2016-04-14" para especificar fecha únicamente	issued	String
Fecha de última actualización/modificación	Recomendado	Fecha de última actualización/modificación del catálogo. Se escribe según el formato ISO-8601, tipeado como fecha simple o fecha con hora, con el mayor detalle posible que sea relevante para el dataset.	"2016-04-19T19:48:05.433640" para especificar fecha y hora "2016-04-19" para especificar fecha únicamente	modified	String
Versión del perfil de metadatos	Recomendado	Es la versión del perfil de metadatos de la red de nodos de datos abiertos de la administración pública nacional de Argentina, utilizada en el catálogo. Se utiliza para que distintas aplicaciones reconozcan y validen los metadatos del catálogo, y las funcionalidades disponibles para distintos fines.	1.1	version	String
Idioma(s)	Recomendado	Lenguaje para la descripción de los metadatos de los datasets contenidos en el catálogo. Hay 2 estándares ISO que pueden ser utilizados para este campo: (a) ISO 639-1 (2 letras) (b) ISO 639-2/T (3 letras) es el más recomendado. Puede definirse 1 o más lenguajes en una lista. (Link a los estándares ISO)	["es"] para un lenguaje ISO 639-1 ["spa", "eng"] para dos lenguajes ISO 639-2	language	Array
Licencia	Recomendado	Indica la licencia bajo la cual todos los datasets y distribuciones del catálogo están disponibles mediante un enlace a la licencia o documento de la licencia seleccionada, o mediante el título textual de la licencia tal como aparece en la lista de http://opendefinition.org/licenses/ . Recomendamos usar la licencia "Open Database License (ODbL) v1.0". Un dataset o distribución que especifique una licencia diferente, sobrescribe a la licencia general del catálogo.	" http://opendatacommons.org/licenses/dbcl/1-0/ " si se utiliza un enlace "Open Database License (ODbL) v1.0" si se consigna el nombre de la licencia a utilizar	license	String
Página web del catálogo	Recomendado	Dirección web de acceso a la página principal del catálogo. Enlace a la página principal del catálogo.	http://datos.gob.ar	homepage	String
Taxonomía global	Sí	Es el sistema de clasificación temática global de la Administración Pública Nacional. Compone una lista de temas globales y está publicada en http://datos.gob.ar/superThemeTaxonomy.json .	http://datos.gob.ar/superThemeTaxonomy.json	superThemeTaxonomy	String
Taxonomía temática específica	Recomendado	Es el sistema de clasificación temática específica, creado por la organización responsable del catálogo. Compone una lista de temas específicos a los datasets del catálogo. Si se clasifica algún dataset del catálogo como perteneciente a uno o más temas, este campo es obligatorio ya que se debe explicitar una taxonomía temática para poder usar sus temas.	[{...}, {...}]	themeTaxonomy	Array

Derechos sobre el catálogo	No	Información sobre derechos aplicables al catálogo en el caso que no se hayan especificado en la licencia. Los datasets y sus distribuciones heredan la información sobre derechos aplicables al catálogo, a menos que especifiquen unos derechos distintos.		rights	String
Cobertura geográfica	No	El área geográfica cubierta por el catálogo. Una región o un lugar determinado. Puede tomar valores: a) De países y, provincias y municipios argentinos, según las recomendaciones de la Guía para la identificación y uso de entidades interoperables. b) Un área de coordenadas representada por latitud/ longitud en el orden: mínima longitud, mínima latitud, máxima longitud, máxima latitud. c) Un punto geográfico representado por latitud/longitud. d) Si la referencia geográfico no está identificada en la Guía para la identificación y uso de entidades interoperables indicar la URIs según geonames.org; ej : http://sws.geonames.org/6255146	"ARG" es el código para la República Argentina. "06007" es el código de un departamento [-58.111111,-35.111111,-57.111111,-33.111111] es un bounding box [-58.111111,-35.111111] es un punto geográfico "http://sws.geonames.org/6255146"	spatial	String or Array

Es importante poner atención a los dos campos que contienen una lista de objetos: **dataset** y **themeTaxonomy**.

El **primero** contendrá una lista de objetos que describen (cada uno) los metadatos de los distintos datasets que componen el catálogo (en la próxima sección se describen los metadatos que debe contener cada uno de estos objetos).

El **segundo** también contiene una lista de objetos que, juntos, definen una taxonomía temática para el catálogo. Cada uno de estos objetos contiene los metadatos que describen a cada uno de los temas de esta taxonomía. Más adelante se describen estos metadatos en la sección Tema.

Dataset (dataset)

A continuación, describimos los metadatos que se deben completar para describir un dataset a la hora de su carga o actualización en el catálogo.

Ejemplos de metadatos de un **dataset**:

- Dataset en texto
- Dataset en JSON

Metadatos que el *data.json* debe contener, para describir a un dataset dentro de la lista contenida en el campo **dataset** del catálogo:

Nombre	Requerido	Descripción	Ejemplo	Variable (data.json)	Tipo (data.json)
Identificador	Sí	Identificador único del dataset, este identificador debe ser único para todo el catálogo.	Un identificador único para el dataset. La URI u otro identificador único en el contexto del catálogo, ejemplo: "dataset-ejemplo-35782"	identifier	String
Título	Sí	Nombre asignado al dataset tal como será publicado. Debe ser claro y lo suficientemente abstracto como para abarcar la multiplicidad de distribuciones que contiene. Se recomienda no exceder los 100 caracteres en la mayoría de los casos. En caso de que un título más largo se juzgue necesario o relevante, este no deberá exceder los 200 caracteres.	Sistema de contrataciones electrónicas	title	String
Descripción	Sí	Descripción del contenido del dataset de un modo claro y conciso. Se recomienda no exceder los 500 caracteres en la mayoría de los casos. En caso de que una descripción más larga se juzgue necesaria o relevante, ésta no deberá exceder los 1500 caracteres.	Datos correspondiente al Sistema de Contrataciones Electrónicas (Argentina Compra)	description	String

Autor	Sí	Responsable de la publicación del dataset. En el caso de organizaciones, detallar la estructura jerárquica separada por puntos, de manera jerárquicamente descendiente. Si la organización es parte de la Administración Pública Nacional y está listada en el dataset llamado "Estructura Organica del Poder Ejecutivo Nacional" (http://datos.gob.ar/dataset/estructura-organica-pen), deberá utilizarse la denominación allí documentada.	Ministerio de Modernización. Secretaría de Modernización Administrativa. Oficina Nacional de Contrataciones.	publisher -> name	String
Correo electrónico del autor	Recomendado	Correo electrónico de contacto del responsable de la publicación del dataset.	onc@modernizacion.gob.ar	publisher -> mbox	String
Área/Persona de contacto	Recomendado	Área/persona de contacto que puede brindar información relevante sobre el dataset.	Ministerio de Modernización. Secretaría de Modernización Administrativa. Oficina Nacional de Contrataciones. Dirección de Compras Electrónicas.	contactPoint -> fn	String
Correo electrónico del área/persona de contacto	Recomendado	Correo electrónico del área/persona de contacto que puede brindar información relevante sobre el dataset.	onc-compraselectronicas@modernizacion.gob.ar	contactPoint -> hasEmail	String
Distribuciones	Sí	Lista de distribuciones que pertenecen al dataset y sus metadatos. Cada distribución se representa con un objeto ("{}") donde se describen los metadatos especificados para la clase "distribution" de este perfil de metadatos.	{...}, {...}	distribution	Array
Temática(s) globales	Sí	Temática/s o categoría/s globales a la/s que se refiere el dataset al ser publicado. Un dataset puede pertenecer a más de una categoría global, de manera que el tipo de valor de este campo es una lista de categorías. La/s categoría/s o temática/s globales deben adoptarse según el campo "Código (authority code)" del Anexo "Taxonomía temática para los datasets", que contiene una lista predefinida de temática/s globales.	["ECON"]	superTheme	Array
Temática(s) específicas	Recomendado	Temática/s o categoría/s específica/s a la/s que se refiere el dataset al ser publicado. Un dataset puede pertenecer a más de una categoría específica, de manera que el tipo de valor de este campo es una lista de categorías. La taxonomía a utilizar debe ser creada por la autoridad responsable del Catálogo. Se deben usar los ids (ver campo "id" de la clase Theme) de los temas definidos en la taxonomía para componer la lista (no se deben usar las etiquetas ni las descripciones).	["Contrataciones", "Compras", "Macroeconomía"]	theme	Array
Fecha de publicación	Sí	Fecha de publicación del dataset. Según el formato ISO-8601, tipeado como fecha simple o fecha con hora, con el mayor detalle posible que sea relevante para el dataset.	"2016-04-14T19:48:05.433640" para especificar fecha y hora "2016-04-14" para especificar fecha únicamente	issued	String
Fecha de última actualización/modificación	Recomendado	Fecha de última actualización/modificación del dataset (ya sea de sus datos o de sus metadatos). Según el formato ISO-8601, tipeado como fecha simple o fecha con hora, con el mayor detalle posible que sea relevante para el dataset.	"2016-04-19T19:48:05.433640" para especificar fecha y hora "2016-04-19" para especificar fecha únicamente	modified	String
Frecuencia de actualización	Sí	Frecuencia con la que se actualiza el dataset. Recomendamos especificar períodos normalizados con formato ISO-8601, agregando el campo "eventual" para datasets que se publican con una frecuencia eventual o no especificada. Anexo "Especificación de frecuencias según ISO-8601".	"R/P1Y" para datasets que se actualizan anualmente	accrualPeriodicity	String
Fuente primaria	No	Fuente original o primaria de los datos publicados en el dataset. Se utiliza cuando la entidad responsable de la publicación del dataset, no es la entidad que produce los datos. En el caso de organizaciones, detallar la estructura jerárquica separada por puntos, de manera jerárquicamente descendiente. Si la organización es parte de la Administración Pública Nacional y está listada en el dataset llamado "Estructura Organica del Poder Ejecutivo Nacional" (http://datos.gob.ar/dataset/estructura-organica-pen), deberá utilizarse la denominación allí documentada.	Ministerio de Hacienda. Instituto Nacional de Estadísticas y Censos. Dirección Nacional de Cuentas Nacionales.	source	String
Página de referencias	No	URL de una página web a través de la cual se puede acceder al dataset, sus recursos o información adicional sobre el mismo.	http://datos.gob.ar/dataset/sistema-de-contrataciones-electronicas-argentina-compra	landingPage	String

Etiqueta(s)	Recomendado	Palabras que describen el título o el contenido del recurso. Es necesario que las etiquetas se encuentren correctamente escritas, en plural y respetando la existencia de tags anteriores. Etiquetas que colaboran en la búsqueda de los usuarios. Cuanto más amplia y uniforme sea la lista de tags mayor será su eficiencia. A tales fines se recomienda ver el Anexo "Pautas para la selección de etiquetas".	["bienes", "compras", "contrataciones"]	keyword	Array
Cobertura temporal	Recomendado	Período de tiempo cubierto por el dataset. El intervalo de tiempo está formado por una fecha de inicio y una de fin separadas por "/", en formato ISO 8601, con el nivel de especificidad requerido por el dataset.	2015-01-01/2015-12-31 2015-01-01T00:45:00Z/2016-01-15T00:06:00Z	temporal	String
Licencia	Recomendado	Indica la licencia bajo la cual el dataset y todas sus distribuciones están disponibles mediante un enlace a la licencia o documento de la licencia seleccionada, o mediante el título textual de la licencia tal como aparece en la lista de http://opendefinition.org/licenses/ . Recomendamos usar la licencia "Open Database License (ODbL) v1.0". Un dataset hereda por default la licencia general del catálogo salvo que se especifique una licencia diferente en este campo. Las distribuciones del dataset heredan esta licencia salvo que especifiquen una diferente.	"http://opendatacommons.org/licenses/dbcl/1-0/" si se utiliza un enlace "Open Database License (ODbL) v1.0" si se consigna el nombre de la licencia a utilizar	license	String
Idioma(s)	No	Lenguaje para la descripción de los metadatos del dataset. Hay 2 estándares ISO que pueden ser utilizados para este campo: (a) ISO 639-1 (2 letras) (b) ISO 639-2/T (3 letras) es el más recomendado. Puede definirse 1 o más lenguajes en una lista. Los lenguajes especificados para un dataset, sobrescriben a los del catálogo. Si este campo está vacío el dataset hereda los lenguajes del catálogo. (Link a los estándares ISO)	["es"] para un lenguaje ISO 639-1 ["spa", "eng"] para dos lenguajes ISO 639-2	language	Array
Cobertura geográfica	No	Una región o lugar determinado al que el dataset haga referencia. Una región o un lugar determinado. Puede tomar valores: a) De países y, provincias y municipios argentinos, según las recomendaciones de la Guía para la identificación y uso de entidades interoperables. b) Un área de coordenadas representada por latitud/ longitud en el orden: mínima longitud, mínima latitud, máxima longitud, máxima latitud. c) Un punto geográfico representado por latitud/longitud. d) Si la referencia geográfica no está identificada en la Guía para la identificación y uso de entidades interoperables indicar la URIs según geonames.org ; ej : http://sws.geonames.org/6255146	"ARG" es el código para la República Argentina. "06007" es el código de un departamento [-58.111111, -35.111111, -57.111111, -33.111111] es un bounding box [-58.111111, -35.111111] es un punto geográfico "http://sws.geonames.org/6255146"	spatial	Array or String

Es importante prestar atención al campo **distribution** que contiene una lista de objetos que describen los metadatos de cada una de las distribuciones del dataset. En la próxima sección abordaremos estos metadatos.

Distribución (distribution)

Estos son los metadatos que se deben completar al cargar o actualizar una distribución de un dataset en el catálogo para describirla.

Ejemplos de metadatos de una **distribución**:

- Distribución en texto
- Distribución en JSON

Metadatos que el *data.json* debe contener, para describir a una distribución dentro de la lista contenida en el campo **distribution** de un dataset:

Nombre	Requerido	Descripción	Ejemplo	Variable (data.json)	Tipo (data.json)
--------	-----------	-------------	---------	----------------------	------------------

Identificador	Sí	Identificador único de la distribución, este identificador debe ser único para la distribución dentro del catálogo completo. Debe estar compuesto por letras mayúsculas o minúsculas de la "a" a la "z" sin caracteres especiales (sin tildes y sin la "ñ"), números, guiones bajos "_", guiones medios "-" y puntos ".".	1.2	identifier	String
Título	Sí	Nombre asignado a la distribución.	Convocatorias abiertas durante el año 2015	title	String
Descripción	Recomendado	Breve descripción de la distribución. Recomendamos no escribir más de una línea. Toda información adicional puede ser incluida en la descripción del dataset.	Listado de las convocatorias abiertas durante el año 2015 en el sistema de contrataciones electrónicas	description	String
URL de descarga	Sí	URL que permite la descarga directa de la distribución del dataset, vincula directamente a un archivo descargable en un formato dado.	http://datos.gob.ar/dataset/becaceb2-dbd0-4879-93bd-5f02bd3b8ca2/resource/bf2f67f4-9ab3-479b-a881-56b43565125e/download/contratos-2015.csv	downloadURL	String
Tipo de distribución	Recomendado	Indica el tipo de recurso. "Archivo de datos" (file): archivo físico de algún formato de datos que se puede descargar. "API" (api): documentación en línea de un servicio web de datos. "Código" (code): repositorio o archivo con scripts utilizados para la generación, transformación, limpieza o validación de los datos de todo o parte del dataset. "Documentación" (documentation): documentación metodológica sobre los datos de todo o parte del dataset.	file api code documentation	type	String
URL de acceso	Sí	URL que permite el acceso a la distribución del dataset. Puede ser una página, feed u otro tipo de recurso que dé acceso indirecto a las distribuciones. Si las distribuciones son solo accesibles a través de la página de referencia del dataset, debe completarse el valor de la URL de acceso a la distribución con el mismo valor de la página de referencia del dataset.	http://datos.gob.ar/dataset/sistema-de-contrataciones-electronicas-argentina-compra/archivo/fa3603b3-0af7-43cc-9da9-90a512217d8a	accessURL	String
Campos de la distribución	Recomendado	Lista de campos que contiene una distribución tabular (no aplica para aquellas distribuciones que no sean tablas) y sus metadatos. Cada campo se representa con un objeto ("{}") donde se describen los metadatos especificados para la clase "field" de este perfil de metadatos (como "nombre", "tipo" y "descripción").	[{...}, {...}]	field	Array
Fecha de publicación	Sí	Fecha de publicación de la distribución. Según el formato ISO-8601, tipeado como fecha simple o fecha con hora, con el mayor detalle posible que sea relevante para el dataset.	"2016-04-14T19:48:05.433640" para especificar fecha y hora "2016-04-14" para especificar fecha únicamente	issued	String
Fecha de última actualización/modificación	Recomendado	Fecha de última actualización/modificación de la distribución. Según el formato ISO-8601, tipeado como fecha simple o fecha con hora, con el mayor detalle posible que sea relevante para el dataset.	"2016-04-19T19:48:05.433640" para especificar fecha y hora "2016-04-19" para especificar fecha únicamente	modified	String
Formato del archivo	Recomendado	Indica el formato del archivo de la distribución. Si el tipo de la distribución está definido por IANA (http://www.iana.org/assignments/media-types/media-types.xml), debe usarse esa definición. En caso contrario deberán ponerse los caracteres después del punto final del archivo, que determinan el formato (cuando no está definido por IANA).	"text/csv" definición de IANA "csv" caracteres finales después del punto	format	String
Nombre del archivo	Recomendado	Nombre de la distribución bajo el cual se descarga un archivo que contiene los datos, incluyendo la extensión del formato. Debe estar compuesto por letras minúsculas de la "a" a la "z" sin caracteres especiales (sin tildes y sin la "ñ"), números y guiones medios "-".	estructura-organica.csv	fileName	String
Licencia	Recomendado	Indica la licencia bajo la cual la distribución está disponible mediante un enlace a la licencia o documento de la licencia seleccionada, o mediante el título textual de la licencia tal como aparece en la lista de http://opendefinition.org/licenses/ . Recomendamos usar la licencia "Open Database License (ODbL) v1.0". Una distribución hereda por default la licencia del dataset al que pertenece, salvo que se especifique una licencia diferente en este campo.	" http://opendatacommons.org/licenses/dbcl/1-0/ " si se utiliza un enlace "Open Database License (ODbL) v1.0" si se consigna el nombre de la licencia a utilizar	license	String
Tipo de archivo	No	Indica el tipo de archivo de la distribución, sólo si este está definido por IANA (http://www.iana.org/assignments/media-types/media-types.xml). En caso contrario este campo permanece vacío.	"text/csv" definición de IANA "" cuando el formato no tiene definición en IANA	mediaType	String
Tamaño	No	Tamaño de la distribución en bytes. El tamaño puede ser aproximado cuando no se conozca el tamaño exacto.	Ejemplo para un archivo de 5Kb aproximadamente: "5120"	byteSize	Integer

Derechos sobre la distribución	No	Información sobre derechos aplicables a la distribución que no se hayan especificado en la licencia. Si se especifican, estos derechos sobreesciben a los del catálogo. En caso contrario, las distribuciones heredan los derechos especificados para el catálogo.		rights	String
--------------------------------	----	--	--	--------	--------

Recomendamos poner atención al campo **field** que contiene una lista de objetos que describen los metadatos de cada uno de los campos de la distribución (en el **caso de distribuciones tabulares, únicamente**). En la próxima sección abordaremos estos metadatos.

Campo (**field**)

Recomendamos enfáticamente que las distribuciones tabulares **incluyan metadatos que ayuden a entender la información que contiene cada campo**. Documentarlos adecuadamente facilita enormemente la correcta utilización de los datos por parte de los usuarios.

En un portal Andino, estos metadatos se completan en el mismo formulario que se utiliza para cargar o actualizar una distribución.

Ejemplos de metadatos de un **campo**:

- Metadatos de un campo de una distribución tabular en JSON

Estos son los metadatos que el *data.json* debe contener para describir a un campo de una distribución tabular dentro de la lista contenida en el campo de metadatos **field** de una distribución:

Nombre	Requerido	Descripción	Ejemplo	Variable (data.json)	Tipo (data.json)
Nombre	Recomendado	<p>El nombre del campo tal como se denomina en el encabezado de la distribución. Véase la "Guía para la publicación de datos en formatos abiertos" para una adecuada elección del nombre de un campo.</p> <p>Se recomienda no exceder los 40 caracteres en la mayoría de los casos. En caso de que un título más largo se juzgue necesario o significativamente más claro, este no deberá exceder los 60 caracteres en ningún caso.</p> <p>Debe estar compuesto por letras minúsculas de la "a" a la "z" sin caracteres especiales (sin tildes y sin la "ñ"), números y guiones bajos "-".</p>	unidad_operativa_contrataciones_desc	title	String
Tipo	Recomendado	<p>El tipo de dato contenido en el campo según la lista utilizada por la librería recline.js (http://okfnlabs.org/recline/docs/models.html#types).</p> <p>Los tipos incluidos en esta lista son:</p> <p>string (text): Valores de texto. number (double, float, numeric): Números que puedan no ser enteros (incluyen decimales). integer (int): Números que siempre son enteros. date: Fecha simple expresada según el estándar ISO 8601 incluyendo únicamente año, mes y día (YYYY-MM-DD) como en "2016-02-01". time: Tiempo expresado según el estándar ISO 8601 incluyendo únicamente horas, minutos y segundos (hh:mm:ss) como en "10:05:00". date-time (datetime, timestamp): Fecha completa expresada según el estándar ISO 8601 incluyendo año, mes, día, horas, minutos y segundos (YYYY-MM-DDThh:mm:ssZ) como en "2016-02-01T10:05:00+03:00". boolean (bool): Valores verdadero/falso. binary: Representación de datos binarios base64. geo_point: Ver estructura en https://www.elastic.co/guide/en/elasticsearch/reference/current/geo-point.html. geojson: ver en http://geojson.org/ array: Lista de valores. object (json): Objeto de JSON. any: Campo que puede contener valores de cualquier tipo.</p>	Convocatorias abiertas durante el año 2015	type	String
Descripción	Recomendado	La descripción de la información que contiene el campo.	Organismo que realiza la convocatoria. Organismo de máximo nivel jerárquico al que pertenece la unidad operativa de contrataciones.	description	String
Identificador	No	<p>El código identificador del campo. Debe ser único para todo el catálogo. Se utiliza cuando el campo requiere un identificador para ser utilizado en un sistema o aplicación, como en el caso de una base de series de tiempo (donde el identificador ejerce el rol de "nomenclador" del campo y debe ser único para todo el sistema - más allá incluso del presente catálogo).</p> <p>Debe estar compuesto por letras mayúsculas o minúsculas de la "a" a la "z" sin caracteres especiales (sin tildes y sin la "ñ"), números, guiones bajos "_", guiones medios "-" y puntos ".".</p>	1.1_OGP_D_1993_A_17	id	String
Unidad de medida	No	La descripción de la unidad de medida en la que están expresados los valores del campo. Sólo se utiliza para campos de tipo numérico.	Millones de pesos a precios de 1993	units	String
Tipo especial	No	<p>El tipo de dato contenido en el campo, correspondiente a alguna extensión especial del perfil de metadatos.</p> <p>El perfil de metadatos contempla el desarrollo de extensiones especiales destinadas a facilitar el desarrollo de aplicaciones automáticas que puedan leer e interpretar el contenido de una distribución o de un campo de un catálogo de datos abiertos, para su uso en sistemas específicos. Estas extensiones especiales buscan brindar mayor información sobre la estructura del dato para mejorar su interoperabilidad desatendida, de una forma sencilla.</p> <p>Para más información leer la sección "Extensiones especiales". Ver "Series de tiempo" como ejemplo de una extensión especial al perfil de metadatos.</p>	time_index	specialType	String
Detalle del tipo especial	No	Un parámetro que se puede especificar según el tipo especial de datos del campo en cuestión. Se utiliza para aportar información adicional sobre la estructura de datos del campo, según su definición en la extensión especial del perfil de metadatos a la que pertenece el "Tipo especial" con el que se define al campo. Sólo se puede usar si antes se definió un "Tipo especial".	R/P1Y	specialTypeDetail	String

Los primeros tres metadatos son útiles para **describir las características de cualquier campo de una distribución tabular**.

Los últimos cuatro metadatos son opcionales porque sólo cobran sentido al **describir las**

características de un campo, para casos específicos. Mientras que no todos los campos de una distribución tabular tienen "**unidad de medida**", la asignación de un "**nomenclador**" o "**identificador**" suele ser útil para la identificación unívoca de variables en otros sistemas o aplicaciones, pero no en la generalidad de los casos.

Préstese especial atención a los últimos dos campos: estos constituyen la forma en que el Perfil de Metadatos **permite el desarrollo de extensiones para sistemas o aplicaciones**, desde su versión 1.1. Ver más en la sección Extensiones especiales.

Tema (theme)

Cada catálogo de datos puede tener su propia taxonomía temática que permite clasificar a los datasets como pertenecientes a una o más categorías temáticas. Recomendamos que los temas tengan algunos metadatos que ayuden a un usuario a entenderlos mejor.

Estos son metadatos que el responsable de cargar o actualizar la taxonomía temática de un catálogo debe completar para describir los temas de la misma.

Ejemplos de metadatos de un **tema**:

- Metadatos de un tema en JSON

Metadatos que el *data.json* debe contener, para describir a un tema de la taxonomía temática de un catálogo:

Nombre	Requerido	Descripción	Ejemplo	Variable (data.json)	Tipo (data.json)
Identificador	Recomendado	El identificador del tema.	AGRI	id	String
Etiqueta	Recomendado	La etiqueta o título de un tema.	Agricultura, pesca, silvicultura y alimentación	label	String
Descripción	Recomendado	Una breve y concisa descripción del tema.	Bajo este concepto se incluyen datasets que cubren dominios tales como agricultura, pesca, forestación o alimentos.	description	String

EXTENSIONES ESPECIALES

A partir de la versión 1.1, el perfil de metadatos propuesto para la Administración Pública Nacional plantea:

1. Un **esquema base de uso general** para la documentación de catálogos de datos abiertos
2. Un marco general para **desarrollar extensiones del perfil**, que permitan documentar casos especiales para el desarrollo de sistemas o aplicaciones.

El desarrollo de extensiones del perfil para uso de aplicaciones puede contemplar:

- La **obligatoriedad de campos de metadatos** que en el perfil base no son obligatorios (recomendados u optativos).
- La definición de uno o más **tipos especiales** (*specialType*) utilizados para que sistemas o aplicaciones interpreten de una forma específica los datos que encuentren en una distribución.

SERIES DE TIEMPO

Junto con la versión 1.1 del perfil de metadatos se propone una sencilla extensión para documentar **distribuciones que contienen series de tiempo**.

Esto sirve para su interpretación y extracción automática por parte de sistemas que compilan series de tiempo en bases de datos, servicios web, y de otras aplicaciones.

Debe tenerse en cuenta que la evolución de las posibilidades soportadas por la definición de esta extensión está directamente relacionada con la evolución de los sistemas desarrollados por la Dirección Nacional de Datos e Información Pública.

Distribución de series de tiempo

Es una tabla donde:

- Existe una columna que es el **índice de tiempo**
- Es la **clave primaria única** de la tabla (identifica una fila única)
- Tiene una **frecuencia temporal determinada** (es diaria, mensual, anual... pero no mezcla

frecuencias).

- Cada una de las otras columnas son **series de tiempo**

indice_tiempo	oferta_global_pib	oferta_global_importacion	demanda_global_exportacion	demanda_global_ibif	demanda_global_consumo_priv
1993-01-01	236520.0336	22027.59999	16340.95975	45069.41348	31952.717
1994-01-01	250307.886	26682.25975	18840.403	51231.4255	32094.804
1995-01-01	243186.1018	24065.62925	23084.79625	44528.27725	32338.89925
1996-01-01	256626.244	28284.11475	24850.043	48483.8615	33040.55475
1997-01-01	277441.3173	35884.496	27876.14225	57047.5	34104.32325
1998-01-01	288123.3068	38903.79175	30837.53425	60780.6695	35249.1645
1999-01-01	278369.0138	34520.59125	30448.89575	53116.3155	36173.34075

La distribución debe ser un archivo CSV estándar (Ver **Guía para la publicación de datos en formatos abiertos**):

- Codificado en `utf-8`
- Separador de campos ,
- Carácter decimal .

```
indice_tiempo,oferta_global_pib,oferta_global_importacion,demanda_global_exportacion,demanda_global_ibif,demanda_global_consumo_priv
1993-01-01,236520.033577,22027.5999938,16340.9597519,45069.4134803,31952.717001
1994-01-01,250307.886,26682.25975,18840.403,51231.4255,32094.804
1995-01-01,243186.10175,24065.62925,23084.79625,44528.27725,32338.89925
1996-01-01,256626.244,28284.11475,24850.043,48483.8615,33040.55475
1997-01-01,277441.31725,35884.496,27876.14225,57047.5,34104.32325
1998-01-01,288123.30675,38903.79175,30837.53425,60780.6695,35249.1645
1999-01-01,278369.01375,34520.59125,30448.89575,53116.3155,36173.34075
```

Ver ejemplo completo de una distribución de series de tiempo.

Precisión de series con números enteros y decimales

Las series deben contener valores con **números enteros** (277441) o con **números decimales** (277441.31725) pero no deben mezclarse dentro de la misma serie, si se desea preservar la precisión. Un número "entero" dentro de una serie decimal debe incluir .0 al final (277441.0).

Los **números enteros** pueden tener cualquier cantidad de cifras, mientras que los **números decimales** sólo pueden tener **hasta 12 cifras en total**, incluyendo la parte entera y la parte decimal:

- 0.12345678901 está ok (*serie decimal*)
- 12345678901.1 está ok (*serie decimal*)
- 1234567.01234 está ok (*serie decimal*)
- 0.1234567890123456789 NO está ok (*serie decimal*)
- 123456789.01234567891 NO está ok (*serie decimal*)
- 1234567890123456789.1 NO está ok (*serie decimal*)
- 123456789123456789123 está ok (*serie entera*)

Casi todos los *software* que manejan números pueden usar decimales con hasta 12 cifras sin pérdida de precisión. Si bien existe software estadístico capaz de manejar decimales con más de 12 cifras, este límite debe ser analizado en cada caso y no está garantizado por la **API de Series de Tiempo de datos.gob.ar**.

Tipo especial: índice de tiempo

Una distribución de series de tiempo puede ser documentada fácilmente en un catálogo especificando uno de sus campos como "índice de tiempo" y aclarando la frecuencia que tiene (Ver el Anexo III - Especificación de frecuencias (según ISO-8601)).

Dentro de la variable `field` de la distribución:

```
json { "specialType": "time_index", "specialTypeDetail": "R/P1Y" }
```

El índice de tiempo de una distribución con series de tiempo debe cumplir:

- **Fechas en ISO 8601** (YYYY-MM-DD)

- **Fechas en orden ascendente** (primera fila tiene el valor más antiguo, última fila el más reciente).
- Fechas pueden tener la parte de "tiempo" luego de la de "fecha", pero se desestima. (YYYY-MM-DD es exactamente igual que YYYY-MM-DDThh:mm:ss).
- Para una frecuencia dada (anual, semestral, trimestral, mensual o diaria) las **fechas admiten usar sólo la parte del estándar de fecha necesaria** o usar la **forma de fecha completa**.
 - Anual: YYYY está ok
 - Anual: YYYY-MM-DD está ok
 - Mensual: YYYY-MM está ok
 - Mensual: YYYY NO está ok
 - Trimestral: YYYY-MM está ok
 - Trimestral: YYYY NO está ok
- Las fechas están completas (el **índice de tiempo es continuo**).
 - Anual: 1980 / 1981 / 1982 está ok
 - Anual: 1980 / 1982 / 1983 NO está ok
 - Diaria
 - Frecuencia diaria *completa* (lunes a domingo). No puede faltar ningún día de la semana.
 - Frecuencia diaria *hábil* o *Business Daily* (lunes a viernes). No puede faltar ningún día hábil de la semana (un "martes" o "lunes" feriado, por ejemplo, debe figurar en el índice y dejar el valor de la observación nulo).
- Para una frecuencia dada (anual, semestral, trimestral, mensual o diaria) donde se use la forma completa, se debe **usar siempre la fecha inicial del período**.
 - Mensual: 1980-01-01 / 1980-02-01 / 1980-03-01 está ok
 - Mensual: 1980-01-31 / 1980-02-28 / 1980-03-31 NO está ok
 - Trimestral: 1980-01-01 / 1980-04-01 / 1980-07-01 / 1980-10-01 está ok
 - Trimestral: 1980-02-01 / 1980-05-01 / 1980-08-01 / 1980-11-01 NO está ok
 - Semestral: 1980-01-01 / 1980-07-01 / 1981-01-01 está ok
 - Semestral: 1980-01-01 / 1980-08-01 / 1981-01-01 NO está ok
 - Semestral: 1980-01-31 / 1980-07-31 / 1981-01-31 NO está ok

Documentar un dataset de series de tiempo

A continuación se revisan los campos del perfil base que adquieren mayor relevancia para documentar series de tiempo, y se desglosa un ejemplo completo para cada parte del modelo de metadatos dentro de un *dataset* que contiene series de tiempo.

Ver ejemplo de catálogo completo en Anexo VII - Ejemplo de data.json con series de tiempo.

- **Dataset 1:** Oferta y Demanda Globales. Datos desestacionalizados. Base 1993
- **Distribucion 1.1:** Oferta y Demanda Global. Precios constantes desestacionalizados. Base 1993. Valores anuales.
- **Distribucion 1.2:** Oferta y Demanda Global. Precios constantes desestacionalizados. Base 1993. Valores trimestrales.

Dataset (dataset) - series de tiempo

```
json { "identifier": "1", "title": "Oferta y Demanda Globales: Datos
desestacionalizados [Base 1993]", "description": "Componentes desestacionalizados
de la oferta y demanda globales a precios de 1993.", "accrualPeriodicity": "R/P3M",
"publisher": { "mbox": "ausolari@mecon.gob.ar", "name": "Ministerio de Hacienda.
Secretaría de Política Económica. Subsecretaría de Programación Macroeconómica." },
"source": "Ministerio de Hacienda. Instituto Nacional de Estadísticas y Censos.
Dirección Nacional de Cuentas Nacionales.", "contactPoint": { "fn": "Ministerio de
Hacienda. Secretaría de Política Económica. Subsecretaría de Programación
Macroeconómica. Dirección de Información y Coyuntura" }, "landingPage":
```



```
datos/blob/master/examples/series_tiempo/distributions/oferta-demanda-global-precios-constantes-desestacionalizados-base-1993-valores-anuales.csv",
"downloadURL": "https://raw.githubusercontent.com/datosgobar/paquete-apertura-datos/master/examples/series_tiempo/distributions/oferta-demanda-global-precios-constantes-desestacionalizados-base-1993-valores-anuales.csv", "field": [ {
"title": "indice_tiempo", "type": "date", "specialType": "time_index",
"specialTypeDetail": "R/PLY" }, { "id": "1.1_OGP_D_1993_A_17", "title":
"oferta_global_pbi", "description": "PIB desestacionalizado, en millones de pesos
de 1993 y valores trimestrales", "type": "number", "units": "Millones de pesos a
precios de 1993" }, { "id": "1.1_OGI_D_1993_A_25", "title":
"oferta_global_importacion", "description": "Importaciones desestacionalizadas, en
millones de pesos de 1993 y valores trimestrales", "type": "number", "units":
"Millones de pesos a precios de 1993" }, { "id": "1.1_DGE_D_1993_A_26", "title":
"demanda_global_exportacion", "description": "Exportaciones desestacionalizadas, en
millones de pesos de 1993 y valores trimestrales", "type": "number", "units":
"Millones de pesos a precios de 1993" }, { "id": "1.1_DGI_D_1993_A_19", "title":
"demanda_global_ibif", "description": "Inversion bruta interna fija
desestacionalizada, en millones de pesos de 1993 y valores trimestrales", "type":
"number", "units": "Millones de pesos a precios de 1993" }, { "id":
"1.1_DGCP_D_1993_A_27", "title": "demanda_global_consumo_priv", "description":
"Consumo privado desestacionalizado, en millones de pesos de 1993 y valores
trimestrales", "type": "number", "units": "Millones de pesos a precios de 1993" }, {
"id": "1.1_DGCP_D_1993_A_30", "title": "demanda_global_consumo_publico",
"description": "Consumo publico desestacionalizado, en millones de pesos de 1993 y
valores trimestrales", "type": "number", "units": "Millones de pesos a precios de
1993" } ] }
```

Campo (field) - series de tiempo

Serie de tiempo

```
json { "id": "1.1_DGCP_D_1993_A_27", "title": "demanda_global_consumo_priv",
"description": "Consumo privado desestacionalizado, en millones de pesos de 1993 y
valores trimestrales", "type": "number", "units": "Millones de pesos a precios de
1993" }
```

Índice de tiempo

```
json { "title": "indice_tiempo", "type": "date", "specialType": "time_index",
"specialTypeDetail": "R/PLY" }
```

GLOSARIO

Ver Glosario

ANEXO I - TAXONOMÍA TEMÁTICA GLOBAL DE LA APN PARA LOS DATASETS (TABLA)

El Portal Nacional de Datos usa la taxonomía temática definida por la Unión Europea.

Cada catálogo de datos puede desarrollar su propia taxonomía, cuyo uso se expresa en los siguientes campos de metadatos (y sus equivalentes para el Portal Andino):

-

themeTaxonomy: es un campo de metadatos del catálogo que define una lista de temas que se pueden usar para clasificar los datasets. Refiere al esquema completo de la taxonomía en sí, no a alguna de sus etiquetas en particular.

-

theme: es un campo de metadatos de un Dataset. Refiere a la/s etiqueta/s en particular bajos la/s cuales un dataset es clasificado temáticamente. Sólo pueden usarse etiquetas que estén definidas en la taxonomía temática de *themeTaxonomy*.

Además del uso de una taxonomía propia de cada catálogo de datos, **recomendamos la clasificación de los datasets según la taxonomía del Portal Nacional de Datos**. Esta es una *súper taxonomía* a la que cada catálogo hace referencia mediante los siguientes campos de metadatos (y sus equivalentes para el Portal Andino):

-

superThemeTaxonomy: es un campo de metadatos del catálogo que apunta a la URL donde el Portal Nacional de Datos documenta la taxonomía temática de la Administración Pública Nacional.

superTheme: es un campo de metadatos de un dataset. Refiere a la/s etiqueta/s en particular bajos la/s cuales un dataset es clasificado temáticamente, según la *súper taxonomía* que es la de la Administración Pública Nacional. Sólo pueden usarse etiquetas que estén definidas en la taxonomía temática de *superThemeTaxonomy*.

La ventaja de usar una súper taxonomía temática es que **facilita la clasificación de datasets** por parte de un usuario según un conjunto de categorías temáticas más generales, que son interoperables con las usadas por otros países del mundo. Esto, a su vez, **facilita la clasificación de datasets cosechados por el Portal Nacional de Datos**.

Etiqueta (label)	Descripción (description)	
AGRI	Agroganadería, pesca y forestación	Datos referidos a agroganadería, pesca y forestación. Por ejemplo: 'Lechería: precio pagado al productor' o 'Superficie forestada'.
ECON	Economía y finanzas	Datos referidos a economía y finanzas. Por ejemplo: 'Deuda pública'.
EDUC	Educación, cultura y deportes	Datos referidos a educación, cultura y deportes. Por ejemplo: 'Registro de Establecimientos Educativos'.
ENER	Energía	Datos referidos a energía. Por ejemplo: 'Productos mineros exportados' o 'Precios del GNC'.
ENVI	Medio ambiente	Datos referidos a medio ambiente. Por ejemplo: 'Operadores de residuos peligrosos'.
GOVE	Gobierno y sector público	Datos referidos a gobierno y sector público. Por ejemplo: 'Inmuebles del estado Nacional'.
HEAL	Salud	Datos referidos a salud. Por ejemplo: 'Estadísticas nacionales de VIH/SIDA'.
INTR	Asuntos internacionales	Datos referidos a asuntos internacionales. Por ejemplo: 'Representaciones argentinas en el exterior'.
JUST	Justicia, seguridad y legales	Datos referidos a justicia, seguridad y legales. Por ejemplo: 'Censo penitenciario'.
REGI	Regiones y ciudades	Datos referidos a regiones y ciudades. Por ejemplo: 'Departamentos de la provincia de Río Negro'.
SOCI	Población y sociedad	Datos referidos a población y sociedad. Por ejemplo: 'Turistas residentes que viajan por Argentina'.
TECH	Ciencia y tecnología	Datos referidos a ciencia y tecnología. Por ejemplo: 'Recursos humanos en ciencia y tecnología'.
TRAN	Transporte	Datos referidos a transporte. Por ejemplo: 'Estadísticas viales'.

ANEXO II - PAUTAS PARA LA SELECCIÓN DE ETIQUETAS

Elegir buenas etiquetas hace más fácil la búsqueda de datasets para los usuarios. Cuanto más amplia y uniforme sea la lista de etiquetas, mayor será su efectividad.

Estas son pautas para definir etiquetas aplicables a la propiedad *keyword* de la clase dataset:

- Escribir correctamente y en plural.
- Usar mayúsculas sólo donde corresponda.
- Identificar palabras claves.
- Respetar la existencia de etiquetas anteriores.
- Agregar sinónimos y emplear lenguaje natural.
- Usar una sólo palabra. Si es muy necesario, usar más de una.
- Si la etiqueta tiene más de una palabra, debe estar separada por un espacio, ej: "declaraciones juradas".

Preguntas útiles a la hora de pensar los etiquetas:

- ¿Cuál es el tema?
- ¿Qué aspectos serán de interés para los usuarios?
- ¿De qué otro modo buscaría sobre esta información?
- ¿De qué tipo de información se trata?
- ¿Qué área la provee?

ANEXO III - ESPECIFICACIÓN DE FRECUENCIAS (SEGÚN ISO-8601)

Frecuencia	Valor según ISO-8601
Cada diez años	R/P10Y
Cada cuatro años	R/P4Y
Cada tres años	R/P3Y
Cada dos años	R/P2Y
Anualmente	R/P1Y
Cada medio año	R/P6M
Cuatrimstralmente	R/P4M
Trimestralmente	R/P3M
Bimestralmente	R/P2M
Mensualmente	R/P1M
Cada 15 días	R/P0.5M
Tres veces por mes	R/P0.33M
Semanalmente	R/P1W
Dos veces a la semana	R/P0.5W
Tres veces a la semana	R/P0.33W
Diariamente	R/P1D
Cada hora	R/PT1H
Continuamente actualizado	R/PT1S
Eventual	eventual

ANEXO IV - EJEMPLO DE DATA.JSON

Este es un ejemplo de data.json:

```
json { "title": "Datos Argentina", "description": "Portal de Datos Abiertos del Gobierno de la República Argentina", "publisher": { "name": "Ministerio de Modernización", "mbox": "datos@modernizacion.gob.ar" }, "issued": "2016-04-14T19:48:05.433640-03:00", "modified": "2016-04-19T19:48:05.433640-03:00", "language": [ "spa" ], "superThemeTaxonomy": "http://datos.gob.ar/superThemeTaxonomy.json", "themeTaxonomy": [ { "id": "convocatorias", "label": "Convocatorias", "description": "Datasets sobre licitaciones en estado de convocatoria." }, { "id": "compras", "label": "Compras", "description": "Datasets sobre compras realizadas." }, { "id": "contrataciones", "label": "Contrataciones", "description": "Datasets sobre contrataciones." }, { "id": "adjudicaciones", "label": "Adjudicaciones", "description": "Datasets sobre licitaciones adjudicadas." }, { "id": "normativa", "label": "Normativa", "description": "Datasets sobre normativa para compras y contrataciones." }, { "id": "proveedores", "label": "Proveedores", "description": "Datasets sobre proveedores del Estado." } ], "license": "Open Data Commons Open Database License 1.0", "homepage": "http://datos.gob.ar", "rights": "Derechos especificados en la licencia.", "spatial": "ARG", "dataset": [ { "title": "Sistema de contrataciones electrónicas", "description": "Datos correspondientes al Sistema de Contrataciones Electrónicas (Argentina Compra)", "publisher": { "name": "Ministerio de Modernización. Secretaría de Modernización Administrativa. Oficina Nacional de Contrataciones", "mbox": "onc@modernizacion.gob.ar" }, "contactPoint": { "fn": "Ministerio de Modernización. Secretaría de Modernización Administrativa. Oficina Nacional de Contrataciones. Dirección de Compras Electrónicas.", "hasEmail": "onc-compraselectronicas@modernizacion.gob.ar" }, "superTheme": [ "ECON" ], "theme": [ "contrataciones", "compras", "convocatorias" ], "keyword": [ "bienes", "compras", "contrataciones" ], "accrualPeriodicity": "R/PLY", "issued": "2016-04-14T19:48:05.433640-03:00", "modified": "2016-04-19T19:48:05.433640-03:00", "identifier": "99db6631-d1c9-470b-a73e-c62daa32c420", "language": [ "spa" ], "spatial": "ARG", "temporal": "2015-01-01/2015-12-31", "landingPage": "http://datos.gob.ar/dataset/sistema-de-contrataciones-electronicas-argentina-compra", "license": "Open Data Commons Open Database License 1.0", "distribution": [ { "accessURL": "http://datos.gob.ar/dataset/sistema-de-contrataciones-electronicas-argentina-compra/archivo/fa3603b3-0af7-43cc-9da9-90a512217d8a", "description": "Listado de las convocatorias abiertas durante el año 2015 en el sistema de contrataciones electrónicas", "format": "CSV", "mediaType": "text/csv", "downloadURL": "http://186.33.211.253/dataset/99db6631-d1c9-470b-a73e-c62daa32c420/resource/4b7447cb-31ff-4352-96c3-589d212e1cc9/download/convocatorias-abiertas-anio-2015.csv", "title": "Convocatorias abiertas durante el año 2015", "license": "Open Data Commons Open Database License 1.0", "byteSize": "5120",
```

```
"issued": "2016-04-14T19:48:05.433640-03:00", "modified": "2016-04-19T19:48:05.433640-03:00", "rights": "Derechos especificados en la licencia.", "field": [ { "title": "procedimiento_id", "type": "integer", "description": "Identificador único del procedimiento de contratación" }, { "title": "organismo_unidad_operativa_contrataciones_id", "type": "integer", "description": "Identificador único del organismo que realiza la convocatoria. Organismo de máximo nivel jerárquico al que pertenece la unidad operativa de contrataciones." }, { "title": "unidad_operativa_contrataciones_id", "type": "integer", "description": "Identificador único de la unidad operativa de contrataciones" }, { "title": "organismo_unidad_operativa_contrataciones_desc", "type": "string", "description": "Organismo que realiza la convocatoria. Organismo de máximo nivel jerárquico al que pertenece la unidad operativa de contrataciones." }, { "title": "unidad_operativa_contrataciones_desc", "type": "string", "description": "Unidad operativa de contrataciones." }, { "title": "tipo_procedimiento_contratacion", "type": "string", "description": "Tipo de procedimiento al que se adecua la contratación." }, { "title": "ejercicio_procedimiento_anio", "type": "date", "description": "Año en el que se inició el proceso de la convocatoria." }, { "title": "fecha_publicacion_convocatoria", "type": "date", "description": "Fecha de publicación de la convocatoria en formato AAAA-MM-DD, ISO 8601." }, { "title": "modalidad_convocatoria", "type": "string", "description": "Modalidad bajo la cual se realiza la convocatoria." }, { "title": "clase_convocatoria", "type": "string", "description": "Clase de la convocatoria." }, { "title": "objeto_convocatoria", "type": "string", "description": "Objeto/objetivo de la convocatoria" } ] ] ] }
```

ANEXO V - TAXONOMÍA TEMÁTICA GLOBAL DE LA APN PARA LOS DATASETS (JSON)

Esta es la taxonomía temática global:

```
json [ { "id": "AGRI", "label": "Agroganadería, pesca y forestación", "description": "Datos referidos a agroganadería, pesca y forestación. Por ejemplo: 'Lechería: precio pagado al productor' o 'Superficie forestada.' }, { "id": "ECON", "label": "Economía y finanzas", "description": "Datos referidos a economía y finanzas. Por ejemplo: 'Deuda pública.' }, { "id": "EDUC", "label": "Educación, cultura y deportes", "description": "Datos referidos a educación, cultura y deportes. Por ejemplo: 'Registro de Establecimientos Educativos.' }, { "id": "ENER", "label": "Energía", "description": "Datos referidos a energía. Por ejemplo: 'Productos mineros exportados' o 'Precios del GNC.' }, { "id": "ENVI", "label": "Medio ambiente", "description": "Datos referidos a medio ambiente. Por ejemplo: 'Operadores de residuos peligrosos.' }, { "id": "GOVE", "label": "Gobierno y sector público", "description": "Datos referidos a gobierno y sector público. Por ejemplo: 'Inmuebles del estado Nacional.' }, { "id": "HEAL", "label": "Salud", "description": "Datos referidos a salud. Por ejemplo: 'Estadísticas nacionales de VIH/SIDA.' }, { "id": "INTR", "label": "Asuntos internacionales", "description": "Datos referidos a asuntos internacionales. Por ejemplo: 'Representaciones argentinas en el exterior.' }, { "id": "JUST", "label": "Justicia, seguridad y legales", "description": "Datos referidos a justicia, seguridad y legales. Por ejemplo: 'Censo penitenciario.' }, { "id": "REGI", "label": "Regiones y ciudades", "description": "Datos referidos a regiones y ciudades. Por ejemplo: 'Departamentos de la provincia de Río Negro.' }, { "id": "SOCI", "label": "Población y sociedad", "description": "Datos referidos a población y sociedad. Por ejemplo: 'Turistas residentes que viajan por Argentina.' }, { "id": "TECH", "label": "Ciencia y tecnología", "description": "Datos referidos a ciencia y tecnología. Por ejemplo: 'Recursos humanos en ciencia y tecnología.' }, { "id": "TRAN", "label": "Transporte", "description": "Datos referidos a transporte. Por ejemplo: 'Estadísticas viales.' } ] }
```

ANEXO VI - EJEMPLO DE METADATOS COMO TEXTO

Este es un ejemplo en markdown:

Catálogo: Datos Argentina

Portal de Datos Abiertos del Gobierno de la República Argentina

-
- Derechos sobre el catálogo: Derechos especificados en la licencia.
-
- Correo electrónico del autor: datos@modernizacion.gob.ar
-

Autor: Ministerio de Modernización

Licencia: Open Data Commons Open Database License 1.0

Idioma(s): spa

Fecha de creación o publicación: 2016-04-14T19:48:05.433640-03:00

Taxonomía temática global: <http://datos.gob.ar/superThemeTaxonomy.json>

Fecha de última actualización/modificación: 2016-04-19T19:48:05.433640-03:00

Cobertura geográfica: ARG

Página web del catálogo: <http://datos.gob.ar>

Taxonomía temática específica

Convocatorias (convocatorias): Datasets sobre licitaciones en estado de convocatoria.

Compras (compras): Datasets sobre compras realizadas.

Contrataciones (contrataciones): Datasets sobre contrataciones.

Adjudicaciones (adjudicaciones): Datasets sobre licitaciones adjudicadas.

Normativa (normativa): Datasets sobre normativa para compras y contrataciones.

Proveedores (proveedores): Datasets sobre proveedores del Estado.

Datasets

Dataset: Sistema de contrataciones electrónicas

Datos correspondientes al Sistema de Contrataciones Electrónicas (Argentina Compra)

Correo electrónico del autor: onc@modernizacion.gob.ar

Autor: Ministerio de Modernización. Secretaría de Modernización Administrativa. Oficina Nacional de Contrataciones

Página de referencias: <http://datos.gob.ar/dataset/sistema-de-contrataciones-electronicas-argentina-compra>

Temática(s) globales: ECON

Fecha de publicación: 2016-04-14T19:48:05.433640-03:00

Cobertura temporal: 2015-01-01/2015-12-31

Fecha de última actualización/ modificación: 2016-04-19T19:48:05.433640-03:00

Idioma(s): spa

- **Temática(s) específicas:** contrataciones, compras, convocatorias
- **Etiqueta(s):** bienes, compras, contrataciones
- **Frecuencia de actualización:** R/PIY
- **Cobertura geográfica:** ARG
- **Licencia:** Open Data Commons Open Database License 1.0
- **Correo electrónico del área/persona de contacto:** onc-compraselectronicas@modernizacion.gob.ar
- **Área/Persona de contacto:** Ministerio de Modernización. Secretaría de Modernización Administrativa. Oficina Nacional de Contrataciones. Dirección de Compras Electrónicas.
- **Identificador:** 99db6631-d1c9-470b-a73e-c62daa32c420

Distribuciones

Distribución: Convocatorias abiertas durante el año 2015

Listado de las convocatorias abiertas durante el año 2015 en el sistema de contrataciones electrónicas

- **URL de acceso:** <http://datos.gob.ar/dataset/sistema-de-contrataciones-electronicas-argentina-compra/archivo/fa3603b3-0af7-43cc-9da9-90a512217d8a>
- **Derechos sobre la distribución:** Derechos especificados en la licencia.
- **Licencia:** Open Data Commons Open Database License 1.0
- **Tamaño:** 5120
- **Formato del archivo:** CSV
- **Tipo de archivo:** text/csv
- **Fecha de última actualización/ modificación:** 2016-04-19T19:48:05.433640-03:00
- **URL de descarga:** <http://datos.gob.ar/dataset/069b5833-e57d-4d7a-859b-67a80cfdf20/resource/fa3603b3-0af7-43cc-9da9-90a512217d8a/download/convocatorias-2015.csv>
- **Fecha de publicación:** 2016-04-14T19:48:05.433640-03:00

Campos de la distribución

- **procedimiento_id** (integer): Identificador único del procedimiento de contratación
- **organismo_unidad_operativa_contrataciones_id** (integer): Identificador único del organismo que realiza la convocatoria. Organismo de máximo nivel jerárquico al que pertenece la unidad operativa de contrataciones.
- **unidad_operativa_contrataciones_id** (integer): Identificador único de la unidad operativa de contrataciones

- **organismo_unidad_operativa_contrataciones_desc** (string): Organismo que realiza la convocatoria. Organismo de máximo nivel jerárquico al que pertenece la unidad operativa de contrataciones.

- **unidad_operativa_contrataciones_desc** (string): Unidad operativa de contrataciones.

- **tipo_procedimiento_contratacion** (string): Tipo de procedimiento al que se adecua la contratación.

- **ejercicio_procedimiento_anio** (date): Año en el que se inició el proceso de la convocatoria.

- **fecha_publicacion_convocatoria** (date): Fecha de publicación de la convocatoria en formato AAAA-MM-DD, ISO 8601.

- **modalidad_convocatoria** (string): Modalidad bajo la cual se realiza la convocatoria.

- **clase_convocatoria** (string): Clase de la convocatoria.

- **objeto_convocatoria** (string): Objeto/objetivo de la convocatoria

ANEXO VII - EJEMPLO DE DATA.JSON CON SERIES DE TIEMPO

Este es un ejemplo de data.json:

```
json { "identifier": "sspm", "title": "Datos Programación Macroeconómica",
"description": "Catálogo de datos abiertos de la Subsecretaría de Programación Macroeconómica.", "publisher": { "name": "Ministerio de Hacienda. Secretaría de Política Económica. Subsecretaría de Programación Macroeconómica.", "mbox": "ausolari@mecon.gob.ar" }, "issued": "2017-09-28T00:00:00", "modified": "2017-09-28T00:00:00", "license": "Open Database License (ODbL) v1.0", "superThemeTaxonomy": "http://datos.gob.ar/superThemeTaxonomy.json", "themeTaxonomy": [ { "id": "nivel_actividad", "description": "Datos sobre nivel actividad", "label": "Nivel actividad" }, { "id": "intercambio_comercial", "description": "Datos sobre intercambio comercial", "label": "Intercambio Comercial" } ], "language": [ "SPA" ], "spatial": "ARG", "dataset": [ { "identifier": "1", "title": "Oferta y Demanda Globales: Datos desestacionalizados [Base 1993]", "description": "Componentes desestacionalizados de la oferta y demanda globales a precios de 1993.", "accrualPeriodicity": "R/P3M", "publisher": { "mbox": "ausolari@mecon.gob.ar", "name": "Ministerio de Hacienda. Secretaría de Política Económica. Subsecretaría de Programación Macroeconómica." }, "source": "Ministerio de Hacienda. Instituto Nacional de Estadísticas y Censos. Dirección Nacional de Cuentas Nacionales.", "contactPoint": { "fn": "Ministerio de Hacienda. Secretaría de Política Económica. Subsecretaría de Programación Macroeconómica. Dirección de Información y Coyuntura" }, "landingPage": "http://www.minhacienda.gob.ar/secretarias/politica-economica/programacion-macroeconomica/", "issued": "2017-08-22T17:51:26.553961-03:00", "keyword": [ "oferta", "demanda", "pbi", "cuentas nacionales", "desestacionalizado" ], "superTheme": [ "ECON" ], "temporal": "1993-01-01/2013-09-30", "theme": [ "nivel_actividad" ], "distribution": [ { "identifier": "1.1", "title": "Oferta y Demanda Globales a precios de 1993: Datos desestacionalizados en valores anuales [Base 1993]", "format": "CSV", "description": "Oferta y Demanda Globales por componente, a precios de comprador, en millones de pesos de 1993 y valores anuales desestacionalizados.", "issued": "2017-08-22T17:51:26.553961-03:00", "modified": "2017-08-22T17:51:26.553961-03:00", "accessURL": "https://github.com/datosgobar/paquete-apertura-datos/blob/master/examples/series_tiempo/distributions/oferta-demanda-global-precios-constant-desestacionalizados-base-1993-valores-anuales.csv", "downloadURL": "https://raw.githubusercontent.com/datosgobar/paquete-apertura-datos/master/examples/series_tiempo/distributions/oferta-demanda-global-precios-constant-desestacionalizados-base-1993-valores-anuales.csv", "field": [ { "title": "indice_tiempo", "type": "date", "specialType": "time_index", "specialTypeDetail": "R/PLY" }, { "id": "1.1_OGP_D_1993_A_17", "title": "oferta_global_pbi", "description": "PIB desestacionalizado, en millones de pesos de 1993 y valores trimestrales", "type": "number", "units": "Millones de pesos a precios de 1993" }, { "id": "1.1_OGI_D_1993_A_25", "title":
```

```
"oferta_global_importacion", "description": "Importaciones desestacionalizadas, en millones de pesos de 1993 y valores trimestrales", "type": "number", "units": "Millones de pesos a precios de 1993" }, { "id": "1.1_DGE_D_1993_A_26", "title": "demanda_global_exportacion", "description": "Exportaciones desestacionalizadas, en millones de pesos de 1993 y valores trimestrales", "type": "number", "units": "Millones de pesos a precios de 1993" }, { "id": "1.1_DGI_D_1993_A_19", "title": "demanda_global_ibif", "description": "Inversion bruta interna fija desestacionalizada, en millones de pesos de 1993 y valores trimestrales", "type": "number", "units": "Millones de pesos a precios de 1993" }, { "id": "1.1_DGCP_D_1993_A_27", "title": "demanda_global_consumo_priv", "description": "Consumo privado desestacionalizado, en millones de pesos de 1993 y valores trimestrales", "type": "number", "units": "Millones de pesos a precios de 1993" }, { "id": "1.1_DGCP_D_1993_A_30", "title": "demanda_global_consumo_publico", "description": "Consumo publico desestacionalizado, en millones de pesos de 1993 y valores trimestrales", "type": "number", "units": "Millones de pesos a precios de 1993" } ] }, { "identifier": "1.2", "title": "Oferta y Demanda Globales a precios de 1993: Datos desestacionalizados en valores trimestrales [Base 1993]", "format": "CSV", "description": "Oferta y Demanda Globales por componente, a precios de comprador, en millones de pesos de 1993 y valores anuales desestacionalizados.", "issued": "2017-08-22T17:51:26.553961-03:00", "modified": "2017-08-22T17:51:26.553961-03:00", "accessURL": "https://github.com/datosgobar/paquete-apertura-datos/blob/master/examples/series_tiempo/distributions/oferta-demanda-global-precios-constant-desestacionalizados-base-1993-valores-trimestrales.csv", "downloadURL": "https://raw.githubusercontent.com/datosgobar/paquete-apertura-datos/master/examples/series_tiempo/distributions/oferta-demanda-global-precios-constant-desestacionalizados-base-1993-valores-trimestrales.csv", "field": [ { "title": "indice tiempo", "type": "date", "specialType": "time_index", "specialTypeDetail": "R/P3M" }, { "id": "1.2_OGP_D_1993_T_17", "title": "oferta_global_pbi", "description": "PBI a precios de comprador, en millones de pesos de 1993 y valores anuales.", "type": "number", "units": "Millones de pesos a precios de 1993" }, { "id": "1.2_OGI_D_1993_T_25", "title": "oferta_global_importacion", "description": "Importación a precios de comprador, en millones de pesos de 1993 y valores anuales.", "type": "number", "units": "Millones de pesos a precios de 1993" }, { "id": "1.2_DGE_D_1993_T_26", "title": "demanda_global_exportacion", "description": "Oferta global total a precios de comprador, en millones de pesos de 1993 y valores anuales.", "type": "number", "units": "Millones de pesos a precios de 1993" }, { "id": "1.2_DGI_D_1993_T_19", "title": "demanda_global_ibif", "description": "Consumo privado, en millones de pesos de 1993 y valores anuales.", "type": "number", "units": "Millones de pesos a precios de 1993" }, { "id": "1.2_DGCP_D_1993_T_27", "title": "demanda_global_consumo_priv", "description": "Consumo publico, en millones de pesos de 1993 y valores anuales.", "type": "number", "units": "Millones de pesos a precios de 1993" }, { "id": "1.2_DGCP_D_1993_T_30", "title": "demanda_global_consumo_publico", "description": "Inversion bruta interna fija, en millones de pesos de 1993 y valores anuales.", "type": "number", "units": "Millones de pesos a precios de 1993" } ] } ] }
```