
Paquete de Apertura de Datos de la República Argentina

Publicación 0.1.0

Datos Argentina

jun. 01, 2017

1. Guía para la publicación de datos en formatos abiertos	1
2. Guía para la identificación y uso de entidades interoperables	13
3. Guía para el uso y la publicación de metadatos	21

Guía para la publicación de datos en formatos abiertos

Índice

- *Introducción*
- *Objetivo de la guía*
- *Formatos abiertos de archivos*
 - *CSV*
 - *JSON*
- *Fragmentación de archivos*
- *Nomenclatura de archivos*
- *Codificación*
- *Estructura y características de los datos tabulares*
 - *Recomendaciones generales*
 - *Nomenclatura de los campos (nombres de las columnas)*
 - *Nivel de granularidad de los datos*
 - *Usar orientación vertical en lugar de horizontal*
 - *Incluir sólo un atributo por campo*
 - *Valores nulos, desconocidos o en blanco en campos numéricos*
 - *Recomendaciones para estructurar planillas de cálculo*
 - *Usar celdas simples*
 - *Fila de encabezado*
 - *Celdas vacías en filas para agrupar conceptos*
 - *Formato de celdas*
 - *Exportación a CSV*
- *Estándares según el tipo de Datos*
 - *Texto*
 - *Entidades*

- *Nombres propios*
- *Siglas*
- *Número*
 - *Moneda*
 - *Números telefónicos*
 - *Coordenadas*
- *Tiempo*
 - *Rangos horarios*
- *Booleano*

Introducción

Esta guía busca ayudar a los organismos a instrumentar la Política de Datos Abiertos impulsada desde el Gobierno de la Nación Argentina, a través del Decreto N° 117/2016 del 12 de enero de 2016.

Objetivo de la guía

Esta es **una guía de buenas prácticas para la publicación de datos en formatos abiertos**.

Estas recomendaciones se basan en:

- Estándares usados a nivel nacional e internacional.
- La experiencia de trabajo del equipo de la Dirección Nacional de Datos e Información Pública del Ministerio 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.

El documento se estructura así:

Formatos abiertos de archivos: cuáles son los formatos más usuales en los que se publican datos y cuáles son los más recomendables.

- **Fragmentación de archivos:** cuáles son los criterios para decidir que un archivo es demasiado grande (y hay que fragmentarlo) o demasiado chico (y hay que juntarlo con otros).
- **Nomenclatura de archivos:** cómo nombrar adecuadamente un archivo.
- **Codificación:** cuál es la codificación en que se debe guardar un archivo.
- **Estructura y características de los datos tabulares**
 - Recomendaciones generales: aplican a todos los casos.
 - Recomendaciones para estructurar planillas de cálculo: aplican exclusivamente al trabajo en planillas de cálculo.
 - Exportación a CSV: cómo exportar adecuadamente planillas de cálculo a un archivo de formato abierto.
- **Estándares según el tipo de datos:** cuáles son los estándares recomendados para manejar distintos tipos de datos.

Estos son los primeros aspectos importantes para la estandarización de datos.

Para una discusión sobre los estándares recomendados en el manejo de datos básicos y fundamentales, transversales a distintas áreas temáticas, se puede consultar la **Guía para la identificación y uso de entidades interoperables**.

Formatos abiertos de archivos

Hay una gran variedad de tecnologías disponibles para producir y almacenar datos. Como ser: planillas de cálculo, bases de datos, software estadístico más específico y más. Esto genera una enorme diversidad de formatos, a veces caótica.

Algunos de estos formatos, no siempre se adecuan a los niveles de apertura deseados. Te ofrecemos algunas pautas y recomendaciones que facilitan la adaptación y/o transformación de estos formatos hacia otros más abiertos y fácilmente reutilizables.

En este cuadro consideramos algunos de los formatos más usados y evaluamos su nivel de apertura:

Antes de seguir, introduciremos dos conceptos que se usarán a lo largo de toda la guía:

- **Distribución o Recurso:** Una distribución o recurso es la unidad mínima en la que se publican datos. Se trata de los archivos que pueden ser descargados y re-utilizados por un usuario. Los recursos pueden tener diversos formatos (.csv, .shp, etc.).
- **Dataset:** Un conjunto de datos o dataset agrupa recursos referidos a un mismo tema que respetan una estructura de la información. Los mismos podrían diferir en el formato en que se los presenta (Ej. .csv, .json, .xls, etc.), la fecha a la que se refieren, el área geográfica cubierta o estar separados bajo algún otro criterio.

Un recurso en formato tabular es un archivo plano que se ajusta a un esquema predefinido de columnas, incluyendo el nombre de la columna y el tipo de datos.

En la mayoría de los casos, corresponde a datos que llegan de bases de datos, reportes y planillas de cálculo en general. A diferencia de los formatos tabulares, los archivos JSON siguen una estructura diferente donde se definen listas de objetos con pares “clave” - “valor”.

Recomendamos con énfasis la publicación de los datos en formato CSV y/o JSON. En caso de utilizar formatos propietarios o aún no estandarizados, es útil indicar software, versión y aplicación que permite procesar esos formatos.

CSV

El CSV es un formato de archivo de texto plano donde las campos (columnas) se separan por comas; y los registros (filas), por saltos de línea. Algunas versiones alternativas de esta forma de publicar datos usan otros separadores como punto y coma (“;”) o pipe (“|”).

Las recomendaciones, siguiendo lo indicado por la especificación [RFC4180](#) y las pautas de la [W3C](#), para la publicación de archivos en formato CSV son:

- Las filas deben finalizar con los caracteres de “retorno de carro” (\r) y “salto de línea” (\n) unidos (\r\n). Esta forma de separar las líneas se denomina usualmente CRLF (“Carriage Return Line Feed”).
- La primera fila siempre contiene los nombres de los campos.
- No se deben repetir nombres entre los campos.
- No se debe colocar espacios al principio ni al final del nombre de un campo, o de un valor.
- Tanto los campos como los valores deben estar separados por comas (“,”).

- En el caso de que un valor contenga el caracter separador (",") o cualquiera de los caracteres que separan las líneas ("\\r", "\\n" o "\\r\\n"), el valor completo debe ser encerrado entre comillas dobles "". Esto indica que el caracter no cumple el rol de separar columnas o filas, sino que es parte de un valor.

Ejemplo:

```
col1,col2\\r\\n
"La tasa de Juan, está vacía",La tasa de Pablo está llena\\r\\n
"La tasa de Juan\\nestá vacía",La tasa de Pablo está llena\\r\\n
"La tasa de Juan\\r\\nestá vacía",La tasa de Pablo está llena\\r\\n
```

- En el caso de que un valor contenga el caracter comilla doble ("), el valor debe ser encerrado entre comillas dobles como en el caso anterior (""") y, además, los caracteres comilla doble que se encuentren dentro del valor deben escribirse dos veces (""").

Ejemplo:

```
col1,col2\\r\\n
"La tasa de ""Juan"" está vacía",La tasa de Pablo está llena\\r\\n
```

- Para todos los tipos de datos se considera válido el valor indefinido. Este se expresará con la ausencia de todo caracter y no con un caracter o string especial como podrían ser "", "null", "none", "nan", etc.

Ejemplo:

```
col1,col2,col3\\r\\n
a,,b\\r\\n
a,"",b\\r\\n
```

JSON

JSON es un formato de texto popular para el intercambio de datos, es un acrónimo de JavaScript Object Notation. Por su característica de ser un formato de tipo estructurado es especialmente útil para el intercambio de datos entre máquina (*machine - readable format*).

El formato JSON ha sido definido por la especificación [RFC 7159](#) y, tal como CSV, también es un estándar abierto.

Fragmentación de archivos

Para garantizar la accesibilidad a los datos, **es necesario fragmentar los archivos excesivamente grandes**, que superen el millón de filas.

Para esto, recomendamos usar conceptos simples, fragmentando:

- por períodos** en caso de tratarse de información temporal (Ej. Años, semestres, trimestres, meses, semanas, días),
- por zonas** en caso de tratarse de información geográfica (Ej. provincias, municipios, barrios, secciones, o manzanas) o
- por dimensiones temáticas** propias del dominio particular de la información.

Sin embargo, siempre que se decida fragmentar un archivo para garantizar su accesibilidad, recomendamos publicar también una versión no fragmentada que contenga el conjunto de datos completo (aunque sea muy grande), a los fines de evitar la tarea de consolidación.

Para eso, sugerimos usar protocolos de compresión, en especial para archivos muy grandes, y altamente compresibles. De hacerlo, aconsejamos protocolos sin pérdida, y abiertos.

Nomenclatura de archivos

Recomendamos estas convenciones para nombrar archivos:

- Usar palabras siempre en minúsculas.
- No usar artículos.
- Usar únicamente letras y números ASCII, siempre en minúsculas, comprendidos en el rango “a-z” y “0-9”.
- Separar las palabras con guión medio “-”.
- En caso de corresponder, ubicar la referencia temporal o del atributo de fragmentación siempre al final.

Ejemplos:

- *acceso-informacion-publica.csv*: Versión completa del recurso.
- *acceso-informacion-publica-2013.csv*: Versión del recurso fragmentada por año.
- *acceso-informacion-publica-201302.csv*: Versión del recurso fragmentada por mes.
- *acceso-informacion-publica-caba.csv*: Versión del recurso fragmentada por división político-territorial (provincia o caba).
- *acceso-informacion-publica-caba-2013.csv*: Versión del recurso fragmentada por división político-territorial (provincia o caba) y año.
- *acceso-informacion-publica-jujuy-20130208.csv*: Versión del recurso fragmentada por división político-territorial (provincia o caba) y fecha.

Para la fragmentación temporal, recomendamos el estándar de los ejemplos, ya que es compacto y ordena los recursos por tiempo: YYYYMMDD. Por favor, recordá mantener siempre dos dígitos para el mes y el día, incluso si el número es menor a 10.

Para la fragmentación por zonas, consultá la **Guía para la identificación y uso de entidades interoperables**, y mirá cómo nombrarlas adecuadamente.

En el caso de usar dimensiones temáticas propias del dominio particular de la información, podés ver esa guía o usar el mejor estándar identificado para esa temática particular.

Codificación

Todos los recursos de datos, incluyendo los geográficos, deben publicarse usando la codificación UTF-8 siguiendo las [recomendaciones de la W3C](#).

Una de las principales razones es que UTF-8 soporta una gran variedad de lenguajes, [según la W3C](#) es un “*estándar en el que se definen todos los caracteres necesarios para la escritura de la mayoría de los idiomas hablados en la actualidad. Su objetivo es ser, y, en gran medida, ya lo ha logrado, un superconjunto de todos los sets de caracteres que se hayan codificado*”.

Estructura y características de los datos tabulares

En esta sección veremos:

- (a) Recomendaciones generales para el trabajo con datos, y
- (b) Recomendaciones para el trabajo con planillas de cálculo, orientadas tanto a facilitar su exportación a formatos abiertos, como a su propia usabilidad en el contexto de cualquier aplicación de planillas de cálculo.

Recomendaciones generales

Estas son recomendaciones generales para el trabajo con datos tabulares. Sugerimos adoptarlas sea cual sea la tecnología usada.

Muchas de las las recomendaciones aquí presentadas se encuadran en los principios de [Tidy Data](#) delineados por [Hadley Wickham](#). Éstos establecen, por ejemplo, que en una tabla de datos “*cada variable es una columna, cada observacion es una fila, y cada tipo de unidad observacional es una tabla*”. Sugerimos complementar la lectura de esta guía con la del trabajo que mencionamos.

Nomenclatura de los campos (nombres de las columnas)

La “nomenclatura de los campos” es el nombre de las columnas en los datos de estructura tabular. Estas recomendaciones aplican a la generalidad de los casos, pero cuando haya convenciones particulares según la temática o rubro de datos de que se trate y éstas entren en conflicto, puede ser conveniente privilegiar primero la convención de la temática específica y luego la convención general.

Los nombres de los campos deben:

- Estar en español.
- Ser lo más explícitos, descriptivos y declarativos como sea posible.
 - Es preferible que el nombre de un campo sea claro antes que corto, pero se recomienda no superar los 50 caracteres.
 - No usar abreviaturas si no es estrictamente necesario o recomendado por una convención ampliamente difundida. En caso de usarlas, incluirlas en el diccionario de datos.
- Estar en minúsculas, no incluir caracteres especiales, ni estar subrayados.
 - Usar palabras compuestas únicamente de caracteres en minúsculas comprendidos en el rango a-z (letras sin tildes) y en el rango 0-9 (dígitos).
 - Las palabras deben unirse con guión bajo “_”.
 - No contener espacios.
 - Las palabras deben separarse siempre con “_”, en lugar de no tener separación alguna: fecha_audiencia_solicitada en lugar de fechaaudienciasolicitada
- Referirse a un sólo atributo de los datos, indivisible en más de un campo.
 - Los campos deben separar los atributos de los datos en la forma más desagregada que sea posible.
 - Se debe evitar definir campos que contengan más de un tipo de información (por ejemplo: e-mail y sitio web, número de teléfono, etc bajo “datos_de_contacto”).
- Si existe una entidad que engloba varias características separadas en campos diferentes, comenzar nombrando los campos con esa entidad y luego con los atributos más específicos (de lo más general a lo más específico).
 - Ej.: solicitante y solicitante_documento son entidades más generales que se repiten en varios campos, que corresponden a atributos más específicos.
- solicitante_nombre
- solicitante_cargo
- solicitante_documento_tipo
- solicitante_documento_numero

- Resulta más fácil identificar qué campos están relacionados entre sí porque configuran atributos de una misma entidad, en lugar de parecer campos conceptualmente independientes. Además, el ordenamiento alfabético de los campos los dejaría automáticamente agrupados por su pertenencia a una entidad más importante.
- Incluso cuando la entidad de un atributo parezca evidente (ej.: un dataset llamado “audiencias” donde todos los campos son atributos de la entidad “audiencia”), se recomienda nombrar el campo incluyendo la entidad a la que hace referencia el atributo.
 - **No recomendado:** “fecha_hora”
 - **Recomendado:** “audiencia_id”, “audiencia_fecha_hora”.

Los campos que sean identificadores o códigos, deberán incluir el sufijo “_id” en el nombre del campo, salvo casos excepcionales donde un nombre alternativo sea más conveniente porque ofrece información sobre el sistema de identificación usado.

- En cuanto a los campos que contengan la descripción de ese identificador, se recomienda que incluyan el sufijo “_desc” (por “descripción”) en caso de que no exista una forma más conveniente de nombrar el campo.

Recomendado

Nivel de granularidad de los datos

Por favor, no incluir totales, subtotaes ni agrupamientos de datos. Un dataset debe ser consistente en el nivel de granularidad de los datos que contiene. Está bien tener un dataset con la cantidad de convenios firmados por provincia y está bien tener un dataset con la cantidad de convenios firmados por municipio. No está bien tener un dataset que mezcle ambos.

Dicho esto, el dato agregado “convenios firmados por provincia” siempre se puede calcular a partir de un proceso del dataset más desagregado, pero esto no es así a la inversa (es imposible recuperar los datos a nivel de municipio desde el dataset provincial).

No recomendado - datos con subtotaes y/o totales incluidos (diferentes niveles de granularidad)

Recomendado - datos de un mismo nivel de granularidad

Usar orientación vertical en lugar de horizontal

Es preferible que la orientación de los datos sea “vertical” en lugar de “horizontal” en los casos en que esto sea posible. La principal razón es que los datos orientados de manera vertical facilitan el tratamiento y análisis de los datos.

No recomendado - Orientación horizontal

Recomendado - Orientación vertical

Incluir sólo un atributo por campo

Se recomienda definir los campos de forma atómica de modo de incluir un sólo atributo por elemento, en lugar de datos múltiples, generando campos adicionales de ser necesario.

No recomendado - múltiples datos en una celda

Recomendado - un dato por celda

Valores nulos, desconocidos o en blanco en campos numéricos

Los valores de los datos deben ser siempre explícitos y respetando el tipo de datos del campo de que se trate. Los elementos o celdas en blanco se interpretarán siempre como “valor ausente”.

Si existen distintas interpretaciones posibles de un “valor ausente”, éstas deben ser explicitadas en un campo aparte. Si sólo hay “valores ausentes” (no hay distintas interpretaciones, es siempre un “valor ausente”) no es necesario agregar una columna adicional.

Es importante destacar, por ejemplo, que cuando un valor numérico sea “0” siempre debe ponerse un “0” como dato (y no un valor nulo, en blanco o vacío).

No recomendado - texto presente en campos numéricos

Recomendado - texto excluido de campos numéricos

Recomendaciones para estructurar planillas de cálculo

Las recomendaciones de esta sección aplican exclusivamente al trabajo sobre planillas de cálculo.

Usar celdas simples

Recomendamos usar celdas simples y, en ningún caso, combinar celdas.

No recomendado - celdas combinadas

Recomendado - celdas simples, sin combinar

Fila de encabezado

Los datos deben contener sólo una fila de encabezado. Desde la segunda fila en adelante, sólo debe haber datos, pero nunca un encabezado.

No recomendado - múltiples filas de encabezado

Recomendado - encabezado de una única fila

Celdas vacías en filas para agrupar conceptos

Recomendamos no dejar celdas vacías en filas bajo la presunción de que valores en blanco posteriores a un valor positivo contienen implícitamente a ese mismo valor en una suerte de “agrupamiento conceptual”.

Este error es muy común en la construcción de planillas de cálculo y suele generar problemas graves cuando cambia el orden original de las filas. Además, impide el uso de tablas dinámicas y otras formas de analizar los datos.

No recomendado - primera celda de la segunda fila vacía

Recomendado - primera celda de la segunda fila completa

Formato de celdas

Las celdas de una planilla de cálculo deben estar formateadas acorde al tipo de datos de que se trate. Específicamente, **los números** siempre deben estar en celdas de formato/tipo “número”, **los campos de tipo textual** deben estar en celdas de formato/tipo “texto” y **los campos de tipo fecha** deben estar en celdas de formato/tipo “fecha”.

Mantener el formato correcto de las celdas según el tipo de datos que contengan:

- Mejora las las probabilidades de que una exportación a otro formato salga correctamente.
- Hace que los datos sean más operables en la propia planilla de cálculo, aprovechando mejor sus funcionalidades.

Exportación a CSV

Insistimos: CSV es el formato más recomendado para la publicación de archivos tabulares. A la hora de exportar una planilla de cálculo a CSV hay 3 parámetros que deben ser especificados durante la exportación, independientemente del software de que se use:

- **Codificación** (encoding, en inglés): siempre debe ser UTF-8.
- **Caracter separador** (separator character, en inglés): siempre debe ser “,” (coma).
- **Caracter calificador** (quote character o enclosing character, en inglés): siempre debe ser ” “ (comillas dobles).

Estándares según el tipo de Datos

El formato recomendado para los distintos tipos de datos está mayormente basado en las especificaciones de la [W3C](#). En los otros casos, las recomendaciones surgen de la experiencia de trabajo del equipo de la Dirección Nacional de Datos e Información Pública y del esfuerzo realizado en la búsqueda de estándares más adecuados.

Texto

- Los campos de texto no deben contener espacios en blanco innecesarios al principio ni al final.

Entidades

Las entidades que aparezcan entre los datos de un campo textual deben tener una descripción única. Es decir, toda mención que se realice a una entidad dada debe hacerse usando exactamente la misma cadena de caracteres cada vez:

- **Las descripciones de entidades** deberían elegirse siempre de forma tal que cumplan con el estándar específico que las describe, en caso de que este exista.
- **Cuando este estándar no existe** y hay dudas respecto del criterio a adoptar para elegir la descripción única de una entidad, debe **privilegiarse siempre aquella que sea lo más explícita**, descriptiva y declarativa posible.

No recomendado

Recomendado

En el ejemplo anterior, los cuatro valores de texto refieren a la misma entidad. Debe elegirse una única forma de referirse a la misma y usarla en todos los casos.

Siempre que sea posible, la elección deberá fundamentarse en el estándar establecido para ese tipo de entidad (para más información ver la [Guía para la identificación y uso de entidades interoperables](#)). En el caso de no existir un estándar, deberá adecuarse a las pautas generales contexto del dataset de que se trate.

Nombres propios

Se capitalizan (primera letra de cada palabra es mayúscula, el resto de las letras son minúsculas) **todas las palabras significativas**, salvo las siglas. Las palabras significativas son aquellas que no cumplen la función de artículos o preposiciones.

Siglas

Todas las siglas se escriben en mayúsculas, sin usar puntos ni espacios intermedios.

Número

- El separador decimal debe ser el caracter ”.”.
- No se usará separador de miles.
- No se deberán usar espacios en blanco.
- Para los números negativos debe incluirse el símbolo menos “-” inmediatamente antes del número, sin espacio en blanco intermedio.

Moneda

Los valores numéricos que sean además valores monetarios se consideran números y, por lo tanto, valen las mismas recomendaciones que para ellos. Además, agregamos las siguientes recomendaciones:

- La cantidad de decimales debe limitarse a dos, salvo que el uso de una mayor cantidad de decimales sea significativo para el caso particular.
- En ningún caso se incluirán símbolos o letras en el campo numérico -ya sea “\$”, “DOL”, “USD”, etc.

Si en el recurso los valores monetarios están expresados en diferentes monedas, se recomienda indicarlo en un campo aparte (que puede llamarse “moneda_id”) usando los códigos alfabéticos definidos en la [ISO 4127](#).

Ejemplo:

- ARS, para el peso argentino.
- USD, para el dólar estadounidense.

Números telefónicos

En este apartado, proponemos una solución para incluir números telefónicos nacionales en los recursos de datos.

A nivel internacional, el estándar para los números telefónicos fue desarrollado por el “Sector de Normalización de las Telecomunicaciones de la Unión Internacional de Telecomunicaciones” (ITU Telecommunication Standardization Sector) bajo la recomendación E.164.

Para el caso de los números nacionales, el ENACOM tiene la competencia sobre el sistema de numeración telefónica. Este organismo determina que el número nacional de abonado debe estar compuesto por 10 dígitos. Estos 10 dígitos están conformados por un indicativo interurbano más un número de abonado. Pudiendo el indicativo interurbano tener entre 2 y 4 dígitos, y el número de abonado entre 6 y 8 dígitos.

Esta numeración es válida para los teléfonos móviles, pero dado que para llamar a un móvil desde un teléfono fijo es necesario anteponer “15” al número de abonado, es necesario que el registro del número telefónico especifique de alguna manera si se trata de un móvil o de un teléfono fijo.

Al no existir estándares nacionales para la inclusión de números telefónicos en recursos de datos, los números telefónicos suelen indicarse de múltiples maneras. Por ejemplo:

No recomendado - Múltiples formas de indicar un número telefónico

Para los recursos que contengan números telefónicos nacionales recomendamos como mínimo:

- Respetar el estándar definido por el Número Nacional Interurbano utilizando la conformación de números mediante 10 dígitos.

- Asegurarse de indicar si el teléfono es móvil o fijo.
- Omitir el agregado de dígitos adicionales en el indicativo interurbano. Recomendamos no indicar cero inicial antes del código de área.

Con las salvedades que comentaremos al final de este apartado, un posible abordaje sería el de la tabla a continuación:

Recomendado - adecuado al estándar del Número Nacional Interurbano

Esta recomendación no contempla estos casos específicos:

- No será aplicable a números de uso público, ejemplo: 100, Bomberos; 911, Policía Federal; etc.
- Para casos que requieran la inclusión de más de un número telefónico. Deberán agregarse campos o modificar la estructura de la base de datos.
- Para teléfonos que requieran la inclusión de un número de interno, y al estar éste definido por la persona u organización específica, deberá considerarse la inclusión de otro campo de tipo texto. Ya que los números de interno pueden incluir texto. Ejemplo: “*86”, “#36”, etc.
- Para el casos de números internacionales recomendamos contemplar el estándar internacional [E.164](#).

Coordenadas

Para registrar datos de coordenadas geográficas de puntos, usamos números decimales. Los campos deberán llamarse “latitud” y “longitud”. Cuando sea conveniente especificar el nombre de la entidad de la cual se consignan las coordenadas, se usarán los sufijos “_latitud” y “_longitud”.

No recomendado

Recomendado

Tiempo

Se usará el estándar [ISO 8601](#) (YYYY-MM-DDTHH:MM:SS[.mmmmmm][+HH:MM]). A menos que se indique lo contrario, se asumirá que la zona horaria es UTC-03:00 (Argentina).

- Fecha: YYYY-MM-DD
- Hora: HH:MM:SS[.mmmmmm][+HH:MM]
- Fecha y Hora: YYYY-MM-DDTHH:MM:SS[.mmmmmm][+HH:MM]
- Duración: YYYY-MM-DDTHH:MM:SS[.mmmmmm]

Rangos horarios

- Los rangos estarán divididos en dos partes separadas por un doble guión bajo “_”, la primera indica el día y la segunda, la hora.
- Se puede omitir la parte del día o bien de la hora pero nunca ambas.
- Si se omite la parte que indica el día se asumirá que el rango abarca todo el horario indicado.
- Si se omite la parte que indica el horario se asumirá que el rango abarca todo el día indicado.
- El día se puede indicar tanto mediante rangos separando los días con guiones medios “-” o bien como particulares con el guión bajo “_”.

Ejemplos de formatos válidos para días:

```
DAY: Un solo día
DAY1-DAY2: Entre entre DAY1 y DAY2
DAY1_DAY2: DAY1 y DAY2
DAY1-DAY2_DAY3: DAY1 a DAY2 y DAY3
```

- La hora se indica mediante rangos, separando los horarios con guiones medios (“-”). También se pueden indicar varios horarios con el guión bajo “_”.

Ejemplos de formatos válidos para horas:

```
HH:MM-HH:MM : Rango simple
HH:MM-HH:MM_HH:MM-HH:MM : Dos rangos
```

Más ejemplos de formatos válidos completos:

```
HH:MM-HH:MM para indicar un rango que ocurre todos los días.
DAY para indicar que el rango ocupa todo el día DAY.
DAY__HH:MM-HH:MM para indicar un rango que ocurre los días DAY entre HH:MM y HH:MM.
DAY__HH:MM-HH:MM_HH:MM-HH:MM para indicar mas un rango horario en el mismo día
DAY1-DAY2__HH:MM-HH:MM para indicar un rango que ocurre los días DAY1 a DAY2 entre HH:
↪MM y HH:MM
DAY1-DAY2__HH:MM-HH:MM_HH:MM-HH:MM para indicar mas un rango horario en el mismo_
↪rango de días
```

- En caso de que se necesite cubrir más de una franja horaria y esta sintaxis sea insuficiente, se pueden incluir varias separadas por espacios.
- Los días se indicarán con sus iniciales en castellano: LUN, MAR, MIE, JUE, VIE, SAB y DOM

Ejemplos:

```
24hs -> "00:00-23:59"
Jueves 24hs -> "JUE"
Jueves de 14:30 a 17 hs -> "JUE__14:30-17:00"
Jueves de 8 a 12 hs y de 16 a 20 hs -> "JUE__08:12-17:00_16:00-20:00"
Jueves de 8 a 15 hs y Viernes de 8 a 15 hs -> "JUE__08:00-15:00 VIE_08:00-15:00"
Lunes a Viernes 7:30 a 17 hs y Sábados 8 a 12 hs -> "LUN-VIE__07:30-17:00 SAB__08:00-
↪12:00"
Lunes a Viernes 8 a 11 y 14 a 18 hs -> "LUN-VIE__08:00-11:00_14:00-18:00"
Lunes y Miercoles 8 a 11 y 14 a 18 hs -> "LUN_MIE__08:00-11:00_14:00-18:00"
Lunes a Miercoles y Viernes 8 a 11 y 14 a 18 hs -> "LUN-MIE_VIE__08:00-11:00_14:00-18:
↪00"
Lunes a Miercoles 8 a 11 y de Viernes a Domingo 9 a 10 -> "LUN-MIE__08:00-11:00 VIE-
↪DOM__09:00-10:00"
```

Booleano

- A menos que se indique lo contrario, se identificarán con los valores *true* o *false*.
 - Esta convención puede variar en algunos rubros específicos de datos, pero en caso de no existir una convención clara y definida aplicable al rubro o contexto del dataset, se recomienda utilizar *true* o *false*.
- Este campo puede contener “valores ausentes”. En ese caso, el campo deberá estar totalmente vacío, no conteniendo ningún carácter.
- Si existe la posibilidad de que haya otro valor que no sea *true*, *false* o “valor ausente” significa que se eligió un tipo de datos incorrecto: este no es booleano, el tipo de dato booleano es binario y sólo admite 2 valores de verdad (aparte del caso del “valor ausente”).

Guía para la identificación y uso de entidades interoperables

Indice

- *Introducción*
- *Objetivo de esta guía*
- *Datos de entidades interoperables*
 - *¿Qué son?*
 - *¿Por qué es importante estandarizarlos?*
- *Tipos de entidades interoperables*
 - *Geográficas*
 - *Países o territorios internacionales*
 - *Divisiones o unidades territoriales internas*
 - ◇ *A. Provincias -> Departamentos -> Fracciones Censales -> Radios Censales (PDFR)*
 - ◇ *B. Provincias -> Municipios (PM)*
 - ◇ *C. Provincias -> Departamentos -> Localidades (PDL)*
 - ◇ *D. Aglomerados*
 - ◇ *E. ¿Cómo nombrar los campos?*
 - *Direcciones y lugares*
 - *Códigos postales*
 - *Personas físicas*
 - *Personas jurídicas*

Introducción

Esta guía busca ayudar a los organismos a instrumentar la Política de Datos Abiertos impulsada desde el Gobierno de la Nación Argentina, a través del Decreto N° 117/2016 del 12 de enero de 2016.

Objetivo de esta guía

Esta es una **guía de buenas prácticas para el uso de entidades interoperables**. Se trata de datos básicos y fundamentales cuyo uso se repite frecuentemente entre datasets de temáticas y fuentes distintas.

Para hacer estas recomendaciones, nos basamos en estándares usados a nivel nacional e internacional y en la experiencia de trabajo del equipo de la Dirección Nacional de Datos e Información Pública del Ministerio 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.

Para una discusión sobre la estandarización de datos, recomendamos consultar la **Guía para la publicación de datos en formatos abiertos**. Este documento se complementa con esa guía y la **Guía para el uso y la publicación de metadatos**.

Datos de entidades interoperables

¿Qué son?

Las entidades interoperables **son aquellas que se repiten y usan frecuentemente dentro de datasets de:**

- **Temáticas diversas entre sí.**
- **Una misma temática** (ej.: Salud), **pero no de otras** (como Educación, Economía, Transporte, etc).

La mayoría de los datasets incluyen campos que responden al dónde, quién, cuándo y qué. Estos campos permiten que los datasets sean interoperables entre sí.

Veamos un ejemplo. Una matriz origen-destino de pasajeros de transporte urbano que dice cuántos viajes se hacen desde la fracción censal A a la fracción censal B, puede interoperar con datos del Censo Nacional sobre las personas que viven en A o en B (desocupación, edad, condiciones de la vivienda, actividad laboral, etc.). Decimos entonces, que la fracción censal es una entidad interoperable.

Algunos ejemplos de entidades interoperables pueden ser:

- **Transversales** (afectan a la mayoría de las áreas temáticas)
 - **¿Dónde?:** geografía (países, provincias, departamentos, fracciones censales, localidades, direcciones, códigos postales).
 - **¿Quién?:** personas (físicas, jurídicas). Entidades (niveles gubernamentales, organismos internacionales, otros países, sociedad civil).
 - **¿Qué?:** categorías presupuesto. Clasificación de bienes transables.
- **Específicas** (afectan a alguna/s área/s temática/s específica/s)
 - **¿Qué?:** actividades económicas. Clasificación de enfermedades. Clasificación de términos clínicos. Clasificación de unidades educativas.

¿Por qué es importante estandarizarlos?

Las entidades interoperables son las que permiten que los datasets hablen entre sí, pero esto no puede suceder cuando dos datasets nombran de forma distinta a una misma entidad interoperable (como cuando se usan distintos sistemas de *ids* o se nombra una misma entidad con/sin mayúsculas, usando artículos y preposiciones (o no usándolos), usando abreviaturas, siglas, tildes, forma corta o completa de un nombre, etc).

Para que los datasets puedan ser interoperables, **deben identificarse todas las entidades interoperables presentes en un dataset y asegurarse de que los datos sobre ellas siguen el mismo estándar.**

A continuación, **proponemos una selección de estándares** producidos por organismos de la Administración Pública Nacional para identificación y uso de entidades interoperables presentes en un activo de datos, en algunas categorías transversales a varias áreas temáticas. **Recomendamos con énfasis usarlos** en todos aquellos casos en los que estén presentes esas entidades. Si por algún motivo esto fuera difícil de aplicar, sugerimos crear un diccionario que permita la traducción de estándares propios a los recomendados.

En los casos de **entidades interoperables específicas sobre alguna temática**, recomendamos **usar el estándar más difundido entre quienes trabajan con frecuencia sobre esa temática dentro de la Administración Pública Nacional.**

Cuando no existan estándares claros dentro de la APN para algún tipo de entidad interoperable en particular, sugerimos **adoptar el mejor estándar internacional en uso**, y seguirlo en forma consistente en todos los datasets que genere el organismo.

La adopción de estándares para el uso de datos de entidades interoperables está **sujeto a cambios y versionados. Por eso, siempre es importante comunicarlos** en forma clara y consistente.

Consideramos a los estándares que recomendamos en este documento como suficientemente estables, abarcativos, difundidos y mantenidos como para que su uso sea beneficioso para el aprovechamiento de los datos y su adopción transparente.

Tipos de entidades interoperables

Geográficas

Países o territorios internacionales

Los nombres y códigos de países o territorios internacionales deben seguir el estándar [ISO 3166-1](#). Recomendamos que el dataset contenga un campo con el código alfabético de 3 dígitos del estándar (Código alfa-3) y otro con el nombre completo del país en español. Para esto, recomendamos usar los “Nombres de uso común” de la [lista de países y sus códigos alfa-3 que publica INDEC](#).

En esta guía, elegimos incluir los nombres de países oficiales y en castellano. Sin embargo, la denominación de los países varía de acuerdo al idioma que se utilice. Por eso, hacemos énfasis en la necesidad de incluir el código de país según el estándar ISO 3166, que es ampliamente usado por organismos internacionales.

A modo de ejemplo, en la Argentina nos referimos a uno de nuestros países vecinos coloquialmente como “Brasil”, mientras que el nombre oficial en portugués es “República Federativa do Brasil” y la traducción oficial en español “República Federativa del Brasil”. El código de país según el estándar definido es “BRA” lo cual resuelve el problema de denominación.

Se recomienda también que el nombre del campo del código sea “pais_id” o, en el caso de que haya más de un campo “país” en el dataset, el nombre de cada campo finalice con “pais_id” (Ej.: “desde_pais_id”, “hasta_pais_id”), mientras que el campo con el nombre completo del país debería ser “pais_nombre”.

No recomendado

Recomendado

Divisiones o unidades territoriales internas

En el caso de las divisiones o unidades territoriales internas, recomendamos usar el sistema de identificadores de la cartografía censal del Censo Nacional 2010 del Instituto Nacional de Estadísticas y Censos ([listado de códigos y](#)

explicación metodológica), que incluye identificadores numéricos compuestos de una cantidad fija de dígitos (el tipo de datos debe ser textual, ya que tiene ceros a la izquierda que son significativos) para, entre otras, las siguientes entidades interoperables:

- Provincias
- Departamentos (Partidos o Comunas)
- Fracciones Censales
- Radios Censales
- Municipios
- Localidades
- Aglomerados

Este sistema de identificadores es consistente con el usado por la Base de Asentamientos Humanos de la República Argentina (BAHRA) para la identificación de localidades, que además lo extiende para incluir la posibilidad de referenciar sitios edificados (sumando 3 dígitos al identificador de una Localidad).

Sin embargo, en esta guía no abordamos los identificadores de entidades puntuales derivadas de las Localidades (que implican avanzar a niveles de desagregación geográfica mayores).

Los nombres geográficos presentados en la BAHRA son oficiales, pero no se encuentran validados. El establecimiento de un procedimiento para la validación de los nombres geográficos es una tarea pendiente para la República Argentina.

¿Cómo se relacionan estas entidades entre sí? Veremos que estas unidades pueden ordenarse jerárquicamente de modo tal que algunas contienen a las otras, aunque no en todos los casos. A continuación, explicamos los conjuntos de entidades que conforman una jerarquía internamente consistente.

A. Provincias -> Departamentos -> Fracciones Censales -> Radios Censales (PDFR)

La provincia es la jurisdicción de primer orden que marca una división político-territorial de la República Argentina. El territorio nacional se divide en 23 de ellas (más la Ciudad de Buenos Aires), siendo que algunos territorios pueden ser marcados como “Indeterminado” (98) o “Sin declarar-Desconocido-Ignorado” (99).

Una **provincia se subdivide** a su vez en jurisdicciones de segundo orden que marcan una división político-administrativa y son llamadas **departamentos** en la mayoría de las provincias (con la excepción de la Provincia de Buenos Aires -Partidos- y la Ciudad de Buenos Aires -comunas-).

Un **departamento, a su vez, se puede subdividir en fracciones censales**, mientras que una **fracción censal se subdivide en radios censales**. Estas son unidades censales que forman parte de la estructura de relevamiento censal.

El tamaño de las fracciones y los radios en áreas urbanas se determina según la cantidad de viviendas. La fracción tiene un promedio de 5000 viviendas mientras que el radio tiene un promedio de 300.

Para bordes de localidades el radio urbano puede bajar a 200 viviendas, aproximadamente, y en localidades aisladas a 100 viviendas. En zonas rurales las fracciones y radios se determinan por la conjunción de distintos factores: características del terreno, accesibilidad y distancia entre las viviendas.

Los identificadores de cada una de estas divisiones se componen, sucesivamente, así:

- **Provincia:** 2 dígitos. Ej.: “06” es la Provincia de “Buenos Aires”.
- **Departamento** (Partido -Provincia de Buenos Aires- o Comuna -Ciudad de Buenos Aires-): 5 dígitos. - Ej.: “06007” es el Departamento “Adolfo Alsina” de la provincia de “Buenos Aires”.
- **Fracciones censales:** 7 dígitos. - Ej.: “0600702” es una Fracción Censal del Departamento “Adolfo Alsina” de la provincia de “Buenos Aires”.

- **Radios censales:** 9 dígitos. - Ej.: “060070201” es un Radio Censal de la Fracción Censal “0600702” del Departamento “Adolfo Alsina” de la provincia de “Buenos Aires”.

B. Provincias -> Municipios (PM)

Los municipios están contenidos por las provincias, pero no las subdividen. Entre medio de ellos puede haber áreas rurales que no pertenezcan a ningún municipio. Los municipios son figuras políticas cuya normativa es potestad de cada provincia y puede haber diferencias significativas entre lo que se considera un municipio en cada una de ellas. Por ejemplo, en algunas provincias los municipios coinciden con los departamentos.

Sin embargo, un municipio siempre está completamente contenido por una sola provincia. Entre las divisiones territoriales internas consideradas en este documento, no hay otra que siempre contenga municipios completos. La superficie de un municipio puede atravesar los límites de departamentos, fracciones y radios censales (un municipio puede estar presente en uno o varios de ellos).

Los identificadores de los municipios se componen entonces con los de las provincias, así:

- **Municipios:** 6 dígitos. - Ej.: “140399” es el Municipio “Camerillo” de la provincia de “Córdoba”.

C. Provincias -> Departamentos -> Localidades (PDL)

Las localidades censales están contenidas tanto por los departamentos como por los municipios. Para componer el identificador se deben usar los departamentos, de tal manera que los primeros 5 dígitos del identificador de una localidad corresponden al identificador del departamento que lo contiene, **los siguientes 3 dígitos son propios de la localidad**:

- **Localidades:** 8 dígitos. - Ej.: “06007010” es la localidad “Carhué” del departamento “Adolfo Alsina” de la provincia de “Buenos Aires”.

La Ciudad de Buenos Aires constituye una excepción a esta regla ya que es una localidad compuesta por departamentos (comunas), de manera que no puede componerse identificador compuesto de tipo provincia-departamento-localidad. Para este caso, recomendamos usar el identificador de jurisdicción de primer nivel de la Ciudad de Buenos Aires (“02”).

D. Aglomerados

Los aglomerados están definidos como conjuntos de localidades y tienen un* id *simple de 4 dígitos (no compuesto) ya que un aglomerado puede cruzar el límite entre 2 municipios, departamentos o provincias.

E. ¿Cómo nombrar los campos?

Al igual que en el caso de los países o territorios internacionales, el dataset debe contener un campo con el código de la división o unidad territorial interna y otro con el nombre o descripción (en caso de que la tenga, anteriormente dijimos que las fracciones y radios censales no tienen nombre o descripción).

Los nombres de los campos identificadores deben ser, respectivamente:

- “provincia_id”
- “departamento_id”
- “fraccion_id”
- “radio_id”

- “municipio_id”
- “localidad_id”
- “aglomerado_id”

Análogamente, debe reemplazarse “_id” por “_nombre” para nombrar los campos que contiene el nombre de cada entidad, cuando esta lo tiene.

Resaltamos la importancia de que el tipo de datos del campo de un identificador es “textual” y no “numérico”. Esto es así porque un valor de tipo numérico no podría comenzar con ceros.

No recomendado

Recomendado

Direcciones y lugares

Siempre que sea posible, cuando un dataset contenga información que identifica a un punto en el espacio geográfico, recomendamos incluir las coordenadas de la manera establecida en la tercera tabla. Las coordenadas de un punto deben ser números decimales negativos o positivos contenidos en dos campos llamados “latitud” y “longitud”.

Si el dataset contiene información sobre direcciones (especialmente en los casos en los que no sea posible proveer coordenadas), recomendamos incluir:

- “calle_nombre”
- “calle_numero”
- “localidad_id”
- “localidad_nombre”
- “provincia_id”
- “provincia_nombre”

Si el dataset incluye direcciones fuera del territorio argentino, deben además incluirse:

- “pais_id”
- “pais_nombre”

No recomendado

Aceptable 1 - sólo dirección normalizada

Aceptable 2 - sólo coordenadas

Recomendado - coordenadas y dirección normalizada

Códigos postales

Los códigos postales deben estar contenidos en un campo llamado “codigo_postal” y seguir el formato definido por el Correo Argentino para el [Código Postal Argentino \(CPA\)](#) a partir de la competencia asignada por la Secretaría de Comunicaciones mediante la Resolución N° 1368/98.

El CPA amplía la información del código postal, incorporando 4 letras que permiten identificar cada cara de manzana en las localidades de más de 500 habitantes. Las localidades de menos de 500 habitantes poseen un único CPA.

El CPA se compone de:

- 1 letra: Identifica a la Provincia.
- 4 números: El Código Postal tradicional.

- 3 letras: Identifican a la Cara de la Manzana.

Ej.: C1426BMD

No recomendado

Recomendado

Personas físicas

Las personas físicas deben identificarse por su nombre completo separado en dos campos (“nombre” y “apellido”), cuando sea posible, donde deben consignarse todos los nombres y todos los apellidos que identifican a un individuo en su documento de identidad oficial, sea el que corresponda según el individuo se presente como residente nacional o extranjero.

Así mismo, recomendamos (de ser posible) agregar dos columnas “id” y “tipo_id” que respectivamente contengan el número o cadena de caracteres que constituye el identificador del documento oficial de la persona y el tipo de documento oficial al que este identificador corresponde (Ej.: DNI, LE, LC y Pasaporte).

Esto es sencillo en el caso de residentes nacionales, pero la variedad de tipos de documentos oficiales que puede presentar un residente extranjero es mucho más amplia y difícil de abarcar. En este último caso es suficiente con consignar si el documento es un “Pasaporte” u “Otro”. Adicionalmente, si el dataset puede contener datos de individuos de diferentes nacionalidades recomendamos agregar un campo “_pais” que contenga la nacionalidad del individuo de referencia.

Tal como explicamos en el caso de países o territorios internacionales, si hubiera más de un campo relativo a “personas” o la mera nomenclatura “nombre” pudiera prestarse a confusión, los campos correspondientes serán compuestos. Ejemplo:

- “sujeto_obligado_nombre”
- “sujeto_obligado_apellido”
- “sujeto_obligado_id”
- “sujeto_obligado_tipo_id”
- “sujeto_obligado_pais_id”
- “sujeto_obligado_pais_nombre”
- “representante_nombre”
- “representante_apellido”
- “representante_id”
- “representante_tipo_id”

No recomendado

Aceptable

Recomendado

Personas jurídicas

Las entidades con personería jurídica local (Ej.: empresas argentinas, ONGs argentinas, etc.) deben registrarse con su CUIT y razón social. Por ejemplo:

- “exportador_cuit”
- “exportador_razon_social”

No recomendado

Recomendado

Si el dataset sólo contiene personas jurídicas registradas en la jurisdicción argentina, el enfoque recomendado para nombrar los campos es el de agregar “_cuit” y “_razon_social” ya que es una nomenclatura específica mucho más descriptiva para el usuario que “_id” y “_desc”. Da cuenta del tipo de “_id” de que se trate y del tipo de descripción asociada.

La Administración Federal de Ingresos Públicos (AFIP) mantiene un [padrón actualizado](#) de todas las personas jurídicas que tienen un CUIT registrado en esa dependencia, que puede ser usado para normalizar o buscar la razón social.

En el caso de que el dataset pueda contener personas jurídicas fuera de la jurisdicción argentina, recomendamos un enfoque análogo al tratamiento de personas físicas:

- “inversor_id”
- “inversor_tipo_id” (Ej.: en el caso de una empresa argentina sería “CUIT”)
- “inversor_desc”
- “inversor_pais_id”
- “inversor_pais_nombre”

Recomendado

Dependiendo de la forma de recolección de los datos, la temática particular del dataset y las capacidades del organismo responsable del mantenimiento del activo de datos, puede ser difícil la recolección comprensible de “_id” y “_tipo_id” de las personas jurídicas de jurisdicción extranjera. Por eso, estos campos pueden llegar a quedar frecuentemente en blanco (valor ausente). Sin embargo, recomendamos con especial énfasis registrar el nombre (“_desc”) de la entidad en cuestión y el país bajo cuya jurisdicción se encuentra.

Guía para el uso y la publicación de metadatos

Indice

- *Glosario*
- *Introducción*
 - *Objetivo de esta guía*
 - *¿Qué son los metadatos?*
 - *¿Cómo se publican los metadatos?*
 - *Público objetivo de esta guía*
- *Portal Andino*
 - *Catálogo*
 - *Dataset*
 - *Distribución*
 - *Campo*
 - *Tema*
- *Otros catálogos*
 - *Estándar usado*
 - *Campos del perfil*
 - *Catálogo*
 - *Dataset*
 - *Distribución*
 - *Campo*
 - *Tema*
- *Anexo I - Taxonomía temática global de la APN para los datasets (tabla)*
- *Anexo II - Pautas para la selección de etiquetas*
- *Anexo III - Especificación de frecuencias (según ISO-8601)*
- *Anexo IV - Ejemplo de data.json*

- *Anexo V - Taxonomía temática global de la APN para los datasets (JSON)*
- *Anexo VI - Ejemplo de metadatos como texto*

Glosario

Activo de datos

Es cualquier **colección de datos con valor informativo**, de propiedad de una entidad u organismo, que reflejan su actividad y son relevantes para el desarrollo de sus misiones y funciones, o para los actores de su ecosistema. Aunque puede designar piezas aisladas de información, suele emplearse para **identificar y tratar conjuntos de datos** que pueden ser comprendidos y tratados **como una única unidad a efectos de su gestión, uso, protección e intercambio**.

Distribución o recurso

Es la unidad mínima de un catálogo de datos. Se trata de los activos de datos que se publican allí y que pueden ser descargados y re-utilizados por un usuario como archivos. Los recursos **pueden tener diversos formatos** (.csv, .shp, etc.). Están acompañados de información contextual asociada (“metadata”) que describe el tipo de información que se publica, el proceso por el cual se obtiene, la descripción de los campos del recurso y cualquier información extra que facilite su interpretación, procesamiento y lectura.

Dataset

También llamado conjunto de datos, es la pieza principal en todo catálogo. Se trata de un activo de datos que agrupa recursos referidos a un mismo tema, que respetan una estructura de la información. Los recursos que lo componen pueden diferir en el formato en que se los presenta (por ejemplo: .csv, .json, .xls, etc.), la fecha a la que se refieren, el área geográfica cubierta o estar separados bajo algún otro criterio.

Catálogo de datos

Es un **directorio de conjuntos de datos, que recopila y organiza metadatos descriptivos**, de los datos que produce una organización. Un portal de datos es un catálogo.

Perfil de metadatos

Es un **estándar para documentar los activos de datos de un catálogo**. En esta guía documentamos un perfil de metadatos que busca unificar el modo en el que estos se comparten desde la Administración Pública Nacional.

Portal Andino

Es el **portal empaquetado y distribuible desarrollado sobre la base de la plataforma CKAN** (portal de datos abiertos), diseñado para facilitar la apertura y federación de activos de datos en la Argentina.

Paquete para la apertura de datos de la República Argentina

Es un **conjunto de herramientas y guías** de buenas prácticas creadas para **hacer más fácil la publicación de datos abiertos** en la Argentina.

Introducción

Esta guía busca ayudar a los organismos a instrumentar la Política de Datos Abiertos impulsada desde el Gobierno de la Nación Argentina, a través del Decreto N° 117/2016 del 12 de enero de 2016.

Objetivo de esta guía

Esta es **una guía de recomendaciones y buenas prácticas, para el uso y la publicación de metadatos**.

Para hacer estas recomendaciones, nos basamos en estándares usados a nivel nacional e internacional y en la experiencia de trabajo del equipo de la Dirección Nacional de Datos e Información Pública del Ministerio 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. Después, existen muchos otros elementos que ayudan al lector a hacer un buen uso de los datos. 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 curada de campos de metadatos, junto con las instrucciones de cómo deben utilizarse, define un **perfil de metadatos**.

En esta guía describimos el **perfil de metadatos recomendado para los catálogos de la Administración Pública Nacional** y cómo publicarlo.

¿Cómo se publican los metadatos?

La publicación de los metadatos puede ser muy diversa en detalle, calidad y forma. **Una publicación muy elemental es un documento de texto** que ofrece una descripción del dataset y de cada uno de los recursos que lo componen. Es posible ver un [ejemplo textual](#) de los metadatos de un catálogo de datos en el **Anexo VI - Ejemplo de metadatos como texto**.

Sin embargo, **las computadoras no pueden leer fácilmente documentos de texto**. La organización sistemática de colecciones de datasets (es decir, la creación de un **catálogo** de datos) exige un nivel de complejidad mayor para facilitar su descubrimiento, indexación, y reutilización por parte de scripts y aplicaciones de todo tipo.

La potencial reutilización de los conjuntos de datos y la posibilidad de que los mismos sean correctamente federados desde el Portal Nacional de Datos Abiertos (datos.gob.ar) dependerá de la calidad de sus metadatos. 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 intenta ayudar a quienes:

- **Publican sus catálogos de datos mediante un adaptación del portal Andino.** Es decir, del portal que el equipo de Datos de la Nación Argentina diseñó sobre la base de CKAN y que pueden adoptar las dependencias de la Administración Pública Nacional (así como también de otros niveles gubernamentales sub-nacionales o incluso entidades del público general) para gestionar la publicación centralizada de sus activos de datos.

El portal Andino incluye formularios web para la carga de datos y metadatos. **La publicación de metadatos mediante este portal cumple con el perfil de metadatos recomendado en esta guía.** 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`

Si estás dentro de este grupo, **recomendamos leer la sección “Portal Andino”**. Con ella, podrás comprender cómo completar cada campo en el formulario web del Portal Andino.

- **Publican sus catálogos de datos mediante otros medios.** Esto es, en forma directa en el dominio digital de la entidad responsable o de alguna forma alternativa en la red. Esas entidades deberán publicar su catálogo de datos en un archivo estructurado (JSON) llamado `data.json` alojado en un dominio de la forma `http://datos.[entidad].gob.ar/data.json` siguiendo las especificaciones del perfil de metadatos de esta guía.

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`).

Si la apertura de tus datos se inscribe en esta modalidad, **te sugerimos leer la sección “Otros catálogos”**. Te ayudará a comprender cómo generar y publicar un archivo `data.json` que cumpla con el perfil de metadatos de la Administración Pública Nacional.

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.**

Al cargar datasets y distribuciones a través de sus formularios, se generan los metadatos del perfil y se publican en un archivo `data.json` en el formato correcto.

El archivo `data.json` de un Portal Andino puede encontrarse en el directorio raíz del portal: `http://datos.[entidad].gob.ar/data.json`.

Catálogo

El perfil de metadatos propuesto se conforma de tres componentes principales:

- La información del catálogo.
- Los datasets.
- Las distribuciones.

El `data.json` de **quienes usen el portal Andino** contará con los metadatos correspondientes a los datasets y distribuciones.

Los metadatos a nivel de catálogo deberán ser completados en una planilla de cálculo y publicados en formato CSV en el directorio raíz del portal, quedando disponibles para su descarga en una URL como *http://datos.[entidad].gob.ar/catalog.csv*.

El resto de los metadatos generados al cargar o actualizar un dataset o una distribución en el portal, se generan a través de los formularios completados y se publican automáticamente en *http://datos.[entidad].gob.ar/data.json*.

Recomendamos ver un [ejemplo textual](#) de metadatos de un **catálogo**.

Estos son los metadatos necesarios para describir el catálogo:

Dataset

Recomendamos ver un [ejemplo textual](#) de metadatos de un **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:

Distribución

Recomendamos ver un [ejemplo textual](#) de metadatos de una **distribución**.

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

Campo

Como se ve en la tabla anterior, metadatos de una distribución, recomendamos que las distribuciones tabulares incluyan algunos metadatos que ayuden a entender la información que contiene cada campo. Estos metadatos se completan en el mismo formulario que se utiliza para cargar o actualizar una distribución.

Estos son los metadatos que el responsable de cargar o actualizar una distribución de un dataset en el catálogo debe completar para describir los campos, en caso de corresponder:

Tema

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:

Otros catálogos

Los usuarios que opten por una solución alternativa 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*.

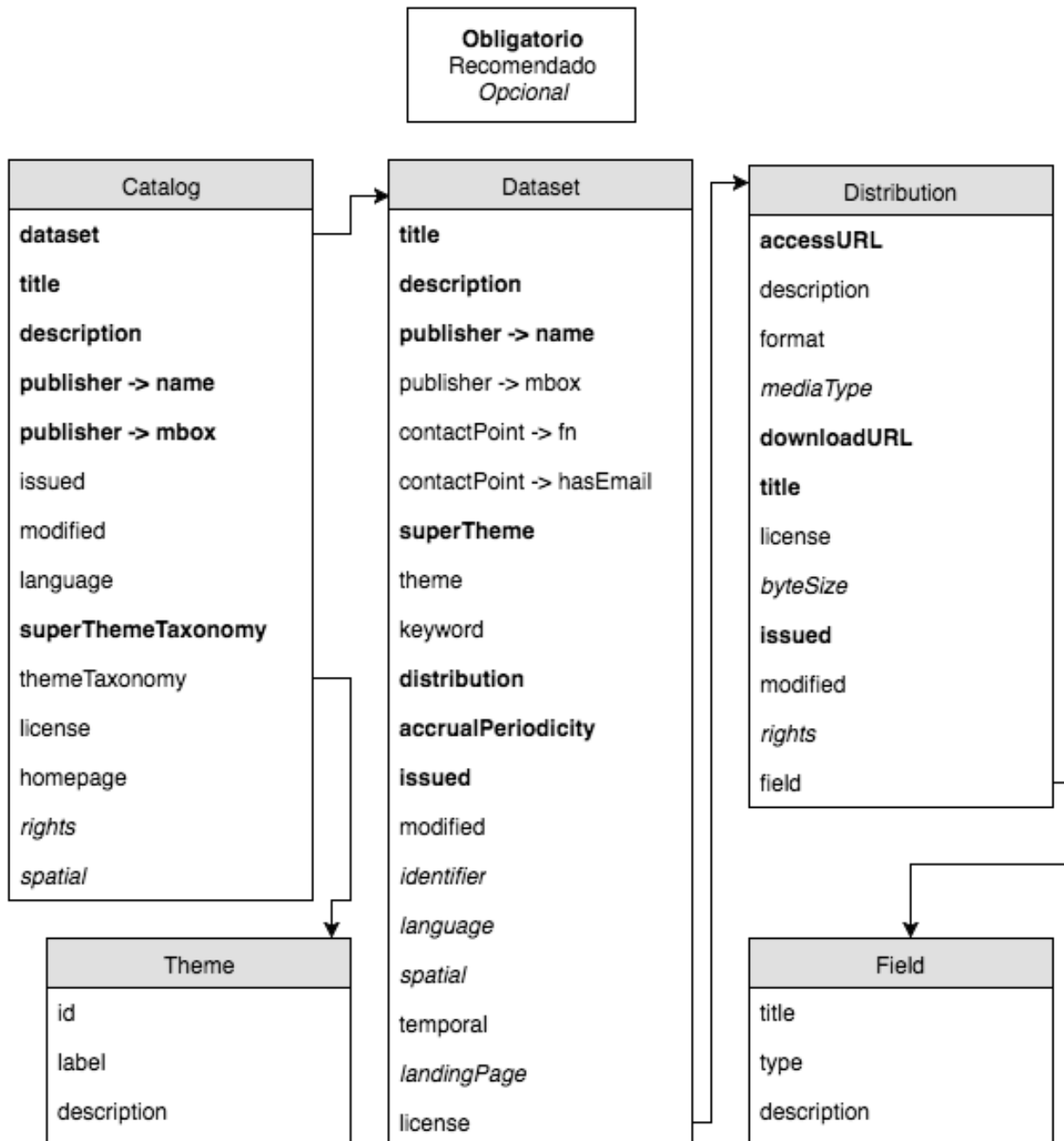
Este archivo deberá estar construido tal como se puede ver en el ejemplo del **Anexo IV - Ejemplo de data.json**, respetando el perfil de metadatos de la Administración Pública Nacional tal como se lo describe más adelante en la sección “*Campos del perfil*”.

Estándar usado

El perfil de metadatos recomendado en esta guía es una adaptación 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

Recomendamos [ver un ejemplo de metadatos de un catálogo en JSON](#).

Estos son los metadatos que el *data.json* debe contener, a nivel de catálogo:

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

Recomendamos [ver un ejemplo de metadatos de un dataset en JSON](#).

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

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

Recomendamos [ver un ejemplo de metadatos de una distribución en JSON](#).

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

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

Recomendamos [ver un ejemplo de metadatos de un campo 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:

Tema

Recomendamos [ver un ejemplo de metadatos de un campo en JSON](#).

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

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 bajo 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 bajo 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**.

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)

Anexo IV - Ejemplo de data.json

Este es un ejemplo de data.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:

```
[
  {
    "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/P1Y
- **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-67a80cfdff20/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