ANEXO III

GUÍA PARA LA IDENTIFICACIÓN Y USO DE ENTIDADES INTEROPERABLES

La última versión de este documento se encuentra en https://datosgobar.github.io/paquete-apertura-datos/guia-interoperables/.

INDICE

- Compras y contrataciones

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 de Datos Públicos de la Jefatura de Gabinete de Ministros 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.: "pais_origen_id", "pais_destino_id"), mientras que el campo con el nombre completo del país debería ser "pais_nombre".

No recomendado

pais_origen	pais_destino	valor_usd
Argentina	China	1405678
República Popular China	argentina	2456786

Recomendado

origen_pais_id	origen_pais_nombre	destino_pais_id	destino_pais_nombre	valor_usd
ARG	Argentina	CHN	China	1405678
CHN	China	ARG	Argentina	2456786

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 Estadística y Censos (ver explicación metodológica).

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 (CSV | SHP | GEOJSON)
- Departamentos (Partidos o Comunas) (CSV | SHP | GEOJSON)
- Fracciones Censales
- Radios Censales
- Municipios (CSV | SHP | GEOJSON)
- Localidades (CSV | JSON)
- Aglomerados

Cabe aclarar que las fracciones censales, la cobertura geográfica, los nomencladores y codificación de INDEC son referencias dinámicas, ya que pueden llegar a modificarse en los censos. Las incluidas en esta guía refieren al Censo Nacional de Población, Hogares y Viviendas 2010.

¿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.

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.

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	Departamento	Fracción Censal	Radio Censal
2 dígitos	5 dígitos	7 dígitos	9 dígitos
"06"	"06007"	"0600702"	"060070201"
Buenos Aires	Adolfo Alsina		

- **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 -> 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** :

Provincia	Departamento	Localidad
2 dígitos	5 dígitos	8 dígitos
"06"	"06007"	"06007010"
Buenos Aires	Adolfo Alsina	Carhué

- **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").

C. 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:

Provincia	Municipio
2 dígitos	6 dígitos
"14"	"140399"
Córdoba	Camerillo

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

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.

Aglomerado
4 dígitos
"0001"
Gran Buenos Aires

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

provincia	flujo_comercial_tipo	valor_usd
Santiago del Estero	Exportación	1405678
Stgo. del Estero	Importación	2456786
Buenos Aires	Exportación	44949874
BA	Importación	44040711

Recomendado

provincia_id	provincia_nombre	flujo_comercial_tipo	valor_usd
86	Santiago del Estero	Exportación	1405678
86	Santiago del Estero	Importación	2456786
06	Buenos Aires	Exportación	44949874
06	Buenos Aires	Importación	440407

F. La Ciudad Autónoma de Buenos Aires

La Ciudad de Buenos Aires constituye una excepción a la regla PDL (provincia-departamento-localidad), utilizada a nivel nacional, 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").

Las comunas son las jurisdicciones de primer orden que marcan la división de la Ciudad Autónoma de Buenos Aires. El territorio municipal se divide en 15 comunas, siendo que algunos territorios pueden ser marcados como "Indeterminado" (98) o "Sin declarar-Desconocido-Ignorado" (99).

Para más detalles sobre el tratamiento de este caso, ver la sección de Divisiones o unidades territoriales internas del GCBA .

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

lugar_nombre	calle_nombre	calle_numero	ciudad
Teatro Colón	Cerrito	604	Ciudad Autónoma de Buenos Aires, CABA

Aceptable 1 - sólo dirección normalizada

lugar_nombre	calle_nombre	calle_numero	localidad_id	localidad_nombre	provincia_id	provincia_nombre
Teatro Colón	Cerrito	604	02001010	Ciudad de Buenos Aires	02	Ciudad Autónoma de Buenos Aires

Aceptable 2 - sólo coordenadas

lugar_nombre	latitud	longitud
Teatro Colón	-34.601041	-58.383079

Recomendado - coordenadas y dirección normalizada

lugar_nombre	calle_nombre	calle_numero	localidad_id	localidad_nombre	provincia_id	provincia_nombre	latitud	longitud
Teatro Colón	Cerrito	604	02001010	Ciudad de Buenos Aires	02	Ciudad Autónoma de Buenos Aires	-34.601041	-58.383079

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

codigo_postal
1426
C 1426 BMD
c1426bmd
C1426

Recomendado

codigo_postal
C1426BMD
C1426BMD
C1426BMD
C1426BMD

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

sujeto_obligado	representante
José Pérez	Carlos Gómez
josé Pérez	Carlos J. Gómez
Pérez, José	Gómez, Carlos
Pérez, José	Gómez, Carlos J.

Aceptable

sujeto_obligado_nombre	sujeto_obligado_apellido	representante_nombre	representante_apellido
José	Pérez	Carlos Jorge	Gómez
José	Pérez	Carlos Jorge	Gómez
José	Pérez	Carlos Jorge	Gómez
José	Pérez	Carlos Jorge	Gómez

Recomendado

sujeto_obligado_nombre	sujeto_obligado_apellido	sujeto_obligado_id	sujeto_obligado_tipo_id	sujeto_obligado_pais_id	sujeto_obligado_pais_nombre
José	Pérez	11111111	DNI	ARG	Argentina
José	Pérez	11111111	DNI	ARG	Argentina
José	Pérez	11111111	DNI	ARG	Argentina
José	Pérez	11111111	DNI	ARG	Argentina

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

exportador
Los Tomates Andinos
Los Tomates
Los Tomates Andinos SRL
Tomates Andinos

Recomendado

exportador_cuit	exportador_razon_social
33111111117	Los Tomates Andinos SRL
33111111117	Los Tomates Andinos SRL
33111111117	Los Tomates Andinos SRL
33111111117	Los Tomates Andinos SRL

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 "_nombre". 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_nombre"
- "inversor_pais_id"
- "inversor_pais_nombre"

Recomendado

inversor_id	inversor_tipo_id	inversor_nombre	inversor_pais_id	inversor_pais_nombre
33111111117	CUIT	Los Tomates Andinos SRL	ARG	Argentina
33111111117	CUIT	Los Tomates Andinos SRL	ARG	Argentina
1234567890	TIN	Tomatoes Inc.	USA	Estados Unidos
987654321	Steuer-Id	Tomate	DEU	Alemania

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 comprehensible 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 ("_nombre") de la entidad en cuestión y el país bajo cuya jurisdicción se encuentra.

ESTÁNDARES SECTORIALES

TRANSPORTE

Se recomienda publicar los datos de movilidad urbana como horarios de transporte público e información geográfica asociada a ellos, según el estándar abierto GTFS (General Transit Feed Standard) .

Los "feeds" GTFS permiten que las empresas de transporte público publiquen sus datos de transporte y que los programadores escriban aplicaciones que consuman esos datos de manera interoperable.

COMPRAS Y CONTRATACIONES

Los datos de compras y contrataciones públicas se publican siguiendo el estándar de Contrataciones Abiertas (Open Contracting Data Standard) . La Argentina se encuentra en proceso de adopción de este estándar desde principios de 2018.