

GUÍA PARA LA APERTURA DE DATOS EN ORGANISMOS DE LA ADMINISTRACIÓN PÚBLICA NACIONAL

INDICE

- Introducción
 - Objetivo
 - Datos abiertos: qué son y por qué tu organismo quiere abrirlos
- 1. Publicá en datos.gob.ar
 - Creá un catálogo de datos abiertos
 - Documentá tus datos
 - Sumá tu catálogo a datos.gob.ar
- 2. Mejorá la calidad de tus datos
 - Compartí tus datos en formatos abiertos
 - Normalizá y enriquecé tus datos
- 3. Publicá datos como un servicio
 - Documentá y compartí tus servicios de datos

- Sumá tus datos a la API de Series de Tiempo
- 4. Identificá nuevos activos de datos a publicar
- ¿Tenés dudas? Escribinos
- Anexos
 - Series de tiempo

INTRODUCCIÓN

Esta guía busca orientar a los organismos a través de los pasos a seguir para 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

El objetivo de esta guía es que los organismos de la Administración Pública Nacional alcanzados por el Decreto N° 117/2016, encuentren todos los recursos necesarios para:

1. Publicar sus datos en datos.gob.ar
2. Mejorar la calidad de sus datos
3. Publicar datos como un servicio
4. Identificar nuevos activos de datos a publicar

DATOS ABIERTOS: QUÉ SON Y POR QUÉ TU ORGANISMO QUIERE ABRIRLOS

Los datos públicos son todos aquellos que el sector público genera y/o administra para el cumplimiento de sus misiones y funciones. No todos los datos públicos *están publicados*; y no es lo mismo *publicar datos* que *abrir datos*.

La diferencia entre un dato *publicado* y uno *abierto* es **todo lo que hace que el dato público sea valioso y fácil de reutilizar** por terceros dentro y fuera del sector público, o incluso por uno mismo:

- Que sea fácil de encontrar.
- Que sea fácil de usar por distintos usuarios, para distintos usos y mediante distintas herramientas.
- Que se pueda cruzar con otros datos (públicos o no, abiertos o no).
- Que se pueda integrar en sistemas y aplicaciones.

Abrir los datos de tu organismo te ayuda a:

- Bajar los costos de compartir datos.
- Organizar tu flujo de trabajo con datos.
- Transparentar la gestión.
- Promover la generación de valor agregado por terceros.

Si querés saber más sobre datos abiertos podés consultar nuestro Kit de Datos Abiertos .

1. PUBLICÁ EN DATOS.GOB.AR

El Portal Nacional de Datos Abiertos (datos.gob.ar) es el punto de acceso para buscar y acceder fácilmente a los datos que publican los organismos de la Administración Pública Nacional.

Si querés incorporar los datos de tu organismo al Portal Nacional, acá te contamos los pasos del proceso para hacerlo:

CREÁ UN CATÁLOGO DE DATOS ABIERTOS

El primer paso para incorporarte a datos.gob.ar es inventariar los datos que tenés y dónde están alojados, y para eso necesitamos que confecciones un catálogo de datos.

Un catálogo de datos abiertos es una lista de activos de datos y sus metadatos, que te describen qué son y cómo podés usarlos. **Facilita encontrar y entender los datos publicados por organismos del Estado** .

Para armar un catálogo, no necesitás tener un portal web;

podés crearlo de varias maneras:

- **Instalando un Portal Andino** . Instalando un portal web de datos abiertos (Ej: <https://datos.agroindustria.gob.ar> - portal de datos abiertos del Ministerio de Agroindustria). Leé más sobre qué es y cómo instalar Portal Andino .
- **Completando un Excel** . Subiendo un archivo Excel a una URL (Ej: <http://estadisticas.produccion.gob.ar/catalogo.xlsx> - catálogo de la Secretaría de Transformación Productiva).
- **Generando un JSON** . Generando y subiendo un archivo JSON a una URL (Ej: <https://www.presupuestoabierto.gob.ar/datasets/data.json> - catálogo de la Subsecretaría de Presupuesto). Esto podrías hacerlo mediante una aplicación propia, otro portal de datos abiertos, etc.

Ver más ejemplos de catálogos .

Leé más sobre cómo crear un catálogo sin usar Portal Andino .

DOCUMENTÁ TUS DATOS

Para que los datos sean abiertos, es fundamental documentarlos como *datasets* en tu catálogo de datos abiertos.

¿Qué datos documento?

Todos los archivos de datos, servicios web de datos y aplicaciones web diseñadas para la descarga de datos que tu organismo tiene en su propiedad digital en internet o en argentina.gob.ar son datos públicos que esperan ser abiertos.

Los archivos XLSX, CSV, DTA, SAV, JSON, SHP y otros formatos tabulares que tu organismo tenga en línea, son activos de datos a documentar en el catálogo.

Las aplicaciones web diseñadas para la consulta y descarga de datos también deben documentarse, como paso previo a generar archivos que permitan descargar el total de la base de datos detrás de la aplicación en lugar de tener que hacer numerosas consultas.

Las APIs de datos deben tener documentación técnica en línea que explique cómo y bajo qué condiciones pueden usarse (se puede generar con Read The Docs u OpenAPI, por ejemplo). En el catálogo de datos abiertos debe registrarse el link a esa documentación técnica.

Tengo datos en archivos PDF

El formato PDF es útil para la publicación de documentos, informes y todo tipo de información, pero **no para la publicación de datos** .

Los datos en archivos PDF no se pueden procesar automáticamente de una manera confiable y estable, obligando al usuario a realizar mucho trabajo manual para estructurar los datos para su uso. Esto hace muy costoso su reutilización y actualización posterior.

Muchas veces estos archivos PDF son generados a partir de tablas en Excel (u otras planillas de cálculo) o bases de datos. ¡En este caso la solución es fácil! Se puede publicar directamente la planilla de cálculo o la base de datos, o incluso implementar un proceso para generar un archivo CSV automáticamente.

Tengo datos en tablas HTML (páginas web)

Las tablas HTML son mejores que los archivos PDF para la publicación de datos, pero aún así no son sencillas de utilizar por la mayoría de los usuarios.

Afortunadamente si tu organismo publica tablas en HTML es muy probable que haya una base de datos, archivo tabular o algún proceso automático detrás. Sólo hay que contactarse con el área que mantiene la página web para que genere y publique la tabla también en un archivo tabular como CSV.

Tengo presentaciones, informes, libros...

Todos estos documentos **no son datos, son información** . Tal vez hubo datos procesados para generarlos, o tal vez están más enfocados en otro tipo de contenido.

En cualquier caso, **no se espera encontrar este tipo de elementos en un catálogo de datos abiertos** . El usuario busca datos que pueda re-procesar por él mismo, para generar contenido o aplicaciones diferentes.

Leé más sobre cómo documentar tus datos en nuestra Guía para el uso y la publicación de metadatos .

El equipo de Datos del Ministerio de Modernización está para ayudarte en este proceso contactanos a datos@modernizacion.gob.ar para que te asistamos en cómo hacerlo mejor y cómo planificar tus próximos pasos.

SUMÁ TU CATÁLOGO A DATOS.GOB.AR

Si tu catálogo cumple con el Perfil Nacional de Metadatos para Datos Abiertos puede integrarse a datos.gob.ar. Cualquier modificación en tu catálogo se verá reflejada automáticamente en datos.gob.ar.

- Escribinos a datos@modernizacion.gob.ar con la URL de tu catálogo para que te sumemos a datos.gob.ar .
- Si sos desarrollador, y querés ver si tu catalogo está listo para estar en datos.gob.ar validalo con **pydatajson** .

A partir de que nos enviás la URL de tu catálogo, el equipo de Datos Argentina analiza cada dataset nuevo que publicás y te ayuda a mejorar la calidad de sus datos y metadatos, antes de agregarlo a datos.gob.ar .

Generalmente recomendamos **trabajar primero en la documentación de 1 dataset del organismo y publicarlo en un primer catálogo de prueba** en una URL pública. Esto nos permite probar cómo funcionaría en datos.gob.ar y **darte una devolución lo antes posible** , facilitando el trabajo de documentar el resto de los datasets.

Ahora que ya estás publicando, te proponemos trabajar en la calidad de tus datos.

2. MEJORÁ LA CALIDAD DE TUS DATOS

Una vez que tenés un catálogo de datos abiertos podés mejorar la calidad de tus datos siguiendo buenas prácticas.

COMPARTÍ TUS DATOS EN FORMATOS ABIERTOS

La diferencia entre un dato *publicado* y uno *abierto* es que este debe ser fácil de utilizar por distintos usuarios, para distintos usos y mediante distintas herramientas.

Leé sobre cómo publicar en formatos abiertos en nuestra Guía para la publicación de datos en formatos abiertos .

Si sos desarrollador y querés implementar rutinas de transformación y limpieza de datos que sigan estas buenas prácticas, podés usar Data Cleaner .

NORMALIZÁ Y ENRIQUECÉ TUS DATOS

Para que tus datos se puedan cruzar fácilmente con otros es necesario que sigas estándares para llamar a las mismas cosas, de la misma manera.

Leé sobre cómo estandarizar tus datos en nuestra Guía para la identificación y uso de entidades interoperables .

Si tenés datos con provincias, departamentos, municipios, localidades, calles o coordenadas podés usar la API del Servicio de Normalización de Datos Geográficos para normalizar o enriquecer entidades geográficas de la Argentina.

3. PUBLICÁ DATOS COMO UN SERVICIO

Publicar datos como un servicio web es un paso más hacia facilitar y potenciar la reutilización de tus datos por parte de analistas, programadores y otros organismos.

Un servicio web es una URL configurable que te permite hacer consultas personalizadas a una base de datos, y actualizarlas fácilmente cuantas veces quieras.

Si ya tenés un servicio de datos, **leé nuestras**

recomendaciones para documentarlo y compartirlo .

Si no tenés un servicio pero publicas series estadísticas, podés **sumarlas a la API de Series de Tiempo de la República Argentina** .

DOCUMENTÁ Y COMPARTÍ TUS SERVICIOS DE DATOS

Si tenés un servicio de datos que querés compartir públicamente:

- Hace que sea más fácil de usar: documentá la interfaz y sus funcionalidades, escribí ejemplos de uso, sé lo más claro posible con los términos de uso. (Ej.: Series de Tiempo , Servicio de Normalización de Datos Geográficos)

Te recomendamos revisar Read the Docs , Mkdocs y OpenAPI como soporte para tu documentación, podés ver nuestro ejemplo de Series de Tiempo y nuestro ejemplo del Servicio de Normalización de Datos Geográficos .

- Ayudá a que sea fácil de encontrar: sumá la URL de la documentación del servicio al catálogo de datos de tu organismo e indicá que es una API en tus metadatos .
- Cuidá tu infraestructura: publicá la base de datos completa en formatos abiertos junto con el servicio (Ej.: dataset de Series de Tiempo).

SUMÁ TUS DATOS A LA API DE SERIES DE TIEMPO

Si publicás indicadores o estadísticas con evolución cronológica (como el nivel de actividad), podés sumarlos fácilmente a la API de Series de Tiempo de la República Argentina y:

- Visualizarlas en el Explorador de Series de Tiempo y compartir tus gráficos.
- Integrar consultas a tus datos en planillas de cálculo .
- Programar consultas a tus datos en distintos lenguajes.

- Armar tableros de seguimiento como este ejemplo usando este proyecto .

Para sumar tus series a la API:

- Publicá tus indicadores o estadísticas en archivos CSV, siguiendo este formato estándar .
- Documentá tus archivos CSV con series de tiempo en tu **catálogo de datos abiertos** .
- Avisanos que tu catálogo tiene series de tiempo , para que lo agreguemos a la API.

Si sos desarrollador, tenés tus series de tiempo en archivos Excel y es difícil convertirlos al formato estándar en CSV, podés usar el Scraper de Series de Tiempo .

Si querés explorar nuevas formas de usar la API de Series de Tiempo para automatizar reportes, análisis, visualizaciones o desarrollar aplicaciones web, mirá este taller .

4. IDENTIFICÁ NUEVOS ACTIVOS DE DATOS A PUBLICAR

Una vez que catalogaste los activos de datos que ya estaban en línea, trabajaste en mejorar su calidad y en mejorar su estructura para facilitar su reutilización, puede ser un buen momento para revisar qué otras cosas podés publicar.

- **Aplicaciones web** . Revisá si hay sistemas o aplicaciones en línea diseñadas para la consulta de datos y si su base de datos (o una consulta procesada sobre ella) puede ser publicada en formatos abiertos.
- **Mapas, visualizaciones, gráficos** . Muchas veces estos se basan en datos, que todavía no están publicados en formatos abiertos. Hay usuarios que quieren sólo consultar el producto final (ie. el mapa), otros que sólo quieren los datos crudos por detrás y a muchos les es útil que publiques ambas cosas. Es una buena idea pensar cualquier nuevo producto digital desde cero con la apertura de sus datos en la hoja de ruta.

- **Estadísticas históricas** . Algunos organismos tienen conjuntos históricos de estadísticas que están "atrapadas" en libros impresos y otros formatos no digitales. Si bien puede ser una tarea ardua, digitalizar y publicar en formatos abiertos estos datos tiene un inmenso valor para la investigación y el análisis con perspectiva histórica.

¿TENÉS DUDAS? ESCRIBINOS

Para comenzar te recomendamos leer nuestro Kit de Datos Abiertos . También podés contactarnos .

ANEXOS

CREAR UN CATÁLOGO EN EXCEL

Para crear un catálogo en Excel:

1. **Descargá este modelo** diseñado para cumplir con el perfil de metadatos.
2. Completá los metadatos de tus datasets, distribuciones y campos.
3. Subí el catálogo a una URL pública y comunicala al equipo de Datos Argentina.
4. Mantené siempre actualizado tu catálogo.

Cada columna del Excel corresponde con un metadato que se verá en datos.gob.ar. Cuando hagas cambios en el Excel, verás los cambios reflejados en datos.gob.ar dentro de las próximas 24hs:

□

El modelo Excel contiene una serie de validaciones para evitar errores en su uso, pero debe tenerse en cuenta que:

- En las hojas de distribución (`distribution`) y campo (`field`) **sólo deben asignarse los ids de los dataset o distribuciones "padre" que los incluyen** , el título se trae automáticamente de la hoja correspondiente con una fórmula.

- **Todas las columnas sean de tipo "texto"** (no "número") salvo aquellas que en el modelo contienen fórmulas para asegurar que una distribución o campo documentado, sean asignados al dataset o distribución correctos.
 - Especialmente las columnas de identificadores (Excel a veces las convierte a número cuando el identificador elegido no contiene letras u otros caracteres).
- **La hoja "referencias" explica el código de colores** para saber qué campos son obligatorios, recomendados, opcionales o avanzados / especiales.
- El modelo contiene comentarios en las celdas para explicar brevemente qué se debe poner en cada una, pero en esta guía se puede consultar la referencia detallada.

SERIES DE TIEMPO

Si tu catálogo es un Portal Andino, revisá la sección Publicar series de tiempo en Andino .

Si tu catálogo es un Excel, revisá la sección Publicar series en la API de Series de Tiempo .

Precisión de series con números enteros y decimales

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

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

- 0.12345678901 está bien (*serie decimal*)
- 12345678901.1 está bien (*serie decimal*)
- 1234567.01234 está bien (*serie decimal*)
- 0.1234567890123456789 NO está bien (*serie decimal*)

- 123456789.01234567891 NO está bien (*serie decimal*)
- 1234567890123456789.1 NO está bien (*serie decimal*)
- 123456789123456789123 está bien (*serie entera*)

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