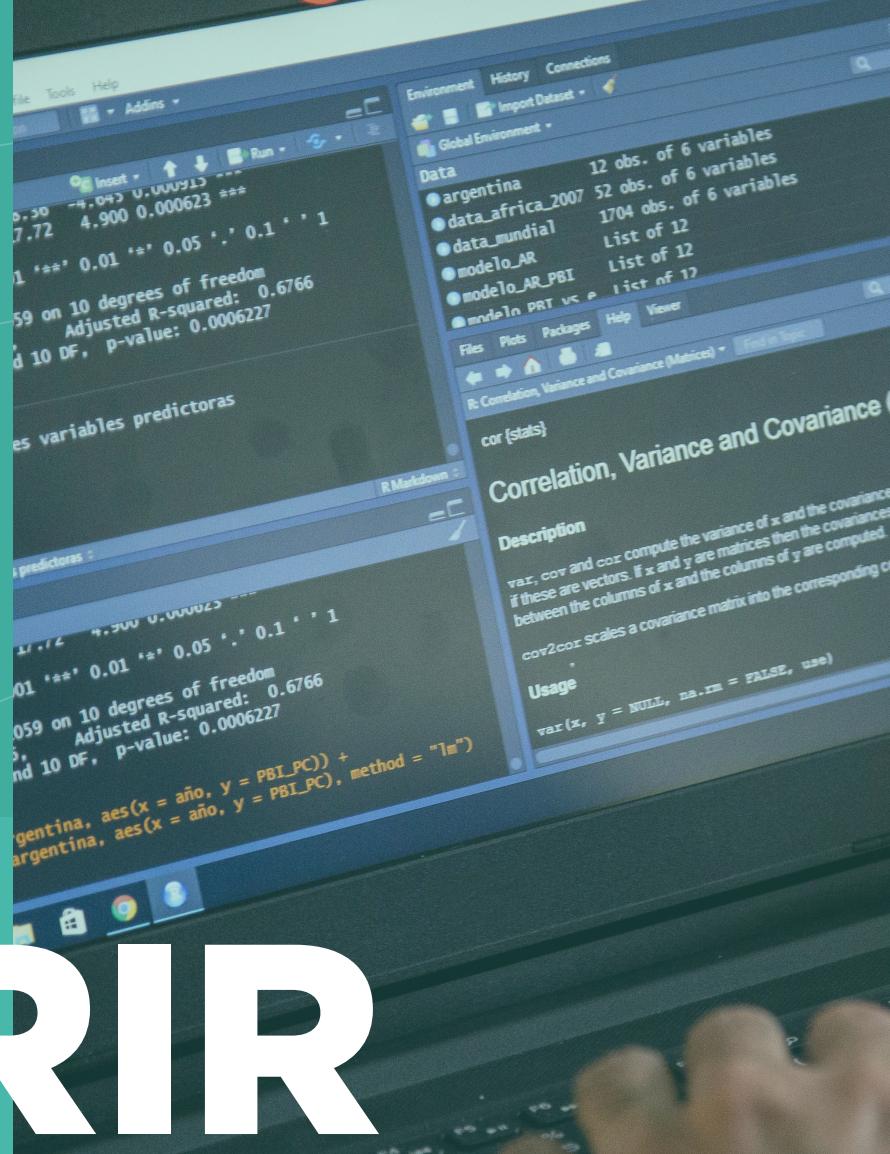


ABRIR DATOS

Claves de una política pública

DATOS.gob.ar



Dirección editorial

Ornella Falcone, Ignacio Leguizamón, Manuel Mercuri, Carolina Raspo.

Investigación y Redacción

Rudi Borrman, Gonzalo Iglesias, Natalia Sampietro, Agustín Benassi, Andrea Miguel, Romina Benítez, Delfina Pérez.

Revisión de contenidos

Lara Purita, Llanes Mariana, Antonella Tártalo.

Impresión

Emilio Etchart, Pablo Conde y Ricardo Gamarra
Imprenta INAP - Secretaría de Gobierno de Modernización



Esta obra está disponible bajo licencia Creative Commons Atribución-NoComercial-CompartirIgual 4.0 Internacional (CC BY-NC-SA 4.0)

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

¿Cómo citar este material?

Academia de Diseño de Políticas Públicas. *Abrir Datos. Claves de una política de apertura* -1era ed. - Buenos Aires. LABgobar, Subsecretaría de Innovación Pública y Gobierno Abierto, Secretaría de Gobierno de Modernización, Argentina, 2019.

www.argentina.gob.ar/datospublicos
www.argentina.gob.ar/academiadediseño

Las consultas o sugerencias sobre los materiales de aprendizaje son bienvenidas y deben dirigirse a:
academia@modernizacion.gob.ar



Este material fue hecho para compartir.
Si ya no lo necesitas, pasáselo a alguien.



La apertura de datos como política pública

Abrir datos es una cuestión global! El proceso de publicar los datos creados, gestionados y/o almacenados por el Estado en formatos abiertos, se convirtió en los últimos años en una política llevada adelante por gran parte de los gobiernos del mundo.

En Argentina, desde la aprobación del Plan de Apertura de Datos (Decreto 117/2016) y la Ley de Acceso a la Información Pública (Nº 27275), cada vez más organismos se sumaron a la iniciativa. Conocé las principales ventajas de esta política:

- **Mejora el proceso de toma de decisiones**

Los datos abiertos permiten diseñar, evaluar y definir los rumbo de las políticas con evidencia. Realizar un ordenamiento, identificación y gestión de la información en la Administración Pública mejora la interoperabilidad y la comunicación interna.

- **Hace posible la transparencia y la rendición de cuentas**

Abir datos permite que los ciudadanos y las organizaciones de la sociedad civil puedan verificar, fiscalizar y monitorear la Administración Pública.

- **Permite crear servicios y aplicaciones**

La posibilidad de reutilizar, relacionar y construir a partir de los datos disponibles genera oportunidades para mejorar y crear nuevas aplicaciones, servicios públicos, proyectos de valor agregado para optimizar la gestión.

- **Contribuye a la innovación y a la generación de valor**

La apertura permite que la ciudadanía, en la medida en que puede conocer y utilizar la información, pueda participar y colaborar activamente en la mejora de los servicios públicos, impulsar la innovación y el desarrollo colectivo de soluciones a problemas comunes. También brinda nuevas posibilidades de negocio al sector privado, a partir de la creación de productos derivados de los datos y la posibilidad de abrir nuevas fuentes de trabajo gracias a los descubrimientos posibilitados por los datos.

Q1

¿Qué son los datos abiertos?

—

¿Cuántas veces leemos una noticia sobre el gasto del Estado en obra pública o usamos una aplicación para definir qué transporte público nos conviene utilizar para llegar a tiempo a un lugar o leemos un artículo académico sobre el impacto de una medida de Gobierno?

Nos informamos y tomamos decisiones sobre nuestras vidas gracias a que las Administraciones Públicas abren datos que producen o almacenan en el ejercicio de sus funciones, permitiendo que otros actores de la sociedad los reutilicen y generen valor agregado a partir de ellos.

a. Los datos, la materia prima

Los datos son elementos descontextualizados, símbolos que describen hechos, condiciones, valores o situaciones. Un dato puede ser una letra, un número o cualquier símbolo que representa una cantidad, una medida, una palabra o una descripción.

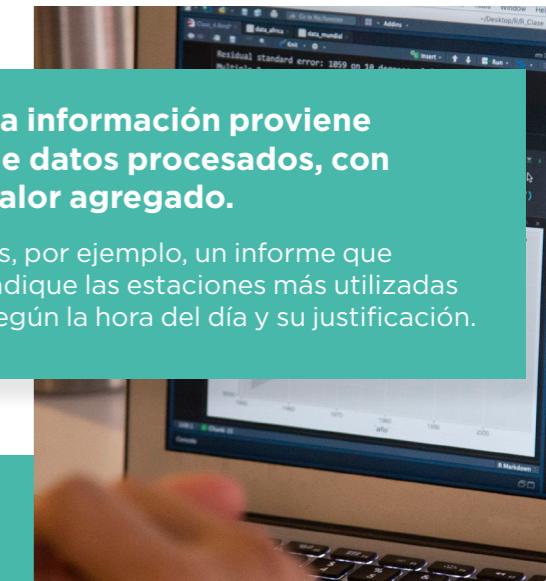
id_provincia	provincia	id_departamento	departamento	sup_sembrada	sup_cosechada	produccion
6	BUENOS AIRES	854	25 DE mayo	3	3	10
6	BUENOS AIRES	882	ZARATE	14	14	50
10	CATAMARCA	7	AMBATO	4	4	4
10	CATAMARCA	112	VALLE VIEJO	5	5	20
22	CHACO	126	1 DE mayo	8	6	24
26	CHUBUT	77	RAWSON	35	35	199
26	CHUBUT	77	RAWSON	22	22	125
14	CORDOBA	168	TOTORAL	22	22	123
14	CORDOBA	168	TOTORAL	20	20	110
18	CORRIENTES	7	BELLA VISTA	1	1	3
18	CORRIENTES	7	BELLA VISTA	3	3	7
30	ENTRE RIOS	113	VILLAGUAY	1	1	3
30	ENTRE RIOS	113	VILLAGUAY	1	1	2
30	ENTRE RIOS	113	VILLAGUAY	1	1	2
34	FORMOSA	14	FORMOSA	0	0	0
38	JIJUY	112	Y...	Estimaciones agrícolas - Secretaría de Gobierno de Agroindustria		

Muchas veces confundimos la publicación de datos con la publicación de información. Si bien ambas publicaciones son relevantes y tienen su uso en cada caso, es importante poder distinguirlas. La información se construye a partir del procesamiento de los datos, que funcionan como mate-

Los datos son elementos descontextualizados.

Los creamos, por ejemplo, cada vez que pasamos la tarjeta por el molinete del subte.

ria prima. Es decir, para obtener información es necesario aplicar algún tipo de transformación o procesamiento sobre los datos, donde se los dota de relevancia y utilidad en busca de generar valor.



b. Acceder, usar y compartir

Los organismos del Estado generan y/o almacenan datos continuamente. Las iniciativas gubernamentales de apertura se concentran principalmente en la publicación de datos públicos en formatos abiertos. Ahora ¿qué diferencia un conjunto de datos públicos de un conjunto de datos abiertos?

Los datos públicos son aquellos que se generan en el ámbito gubernamental, o que se encuentran bajo su guarda. No todos pueden o deben publicarse, algunos están protegidos por legislación específica que los prohíbe o regula.

Los datos abiertos son datos públicos accesibles, disponibilizados en un medio digital en formatos reutilizables siguiendo estándares y buenas prácticas, que usan explícitamente una licencia libre.

- **Accesibilidad:** los datos abiertos deberían estar puestos a disposición de manera visible y gratuita, sin barreras burocráticas o administrativas que puedan disuadir a las personas de acceder a ellos.

- **Formatos reutilizables:** los datos deben estar en formatos que faciliten su uso y reutilización, para que los usuarios puedan trabajar la información que contienen. Diferentes áreas temáticas cuentan con estándares y guías de buenas prácticas locales, nacionales e incluso internacionales. Entre los procedimientos recomendables es importante considerar los metadatos, es decir, grupos de datos que describen el contenido informativo de un recurso o datasets. Metadatos como la fecha de creación, qué organismo es responsable de la base, la versión de la base datos, son útiles para poder contextualizar y entender el origen de los datos para el análisis y uso por parte de terceros.
- **Licencia libre:** es importante la adhesión de forma explícita a una licencia que permita la reutilización de los datos sin restricciones, evitando limitaciones al uso comercial. Recomendamos usar la licencia **Creative Commons Attribution 4.0**, que permite compartir y redistribuir el material en cualquier medio o formato, así como adaptar, remezclar y construir a partir del material, en tanto se cumpla con la atribución correspondiente, citando las fuentes e indicando si se han realizado cambios.

c. ¿Quiénes los usan?

Uno de los aspectos fundamentales que determinan la sostenibilidad de una política de datos abiertos es la identificación y el desarrollo de una comunidad de actores que los reutilicen. Es importante saber quiénes son y de qué maneras pueden aprovechar la información para generar valor con ella.

Sus perspectivas, como usuarios finales, enriquecen el proceso de creación de la política de apertura y permiten que surjan nuevos y/o mejores productos y servicios.



- **Los funcionarios y empleados públicos.** Son un segmento muy relevante de la comunidad de usuarios y consumidores de los datos abiertos. Tener un portal que concentre diferentes activos de datos facilita su labor diaria y a la vez les permite seguir agregando valor en la gestión pública.
- **Organizaciones civiles.** Son catalizadoras de las demandas sociales, cuentan con agendas de incidencia que le permite a la iniciativa de datos publicar aquellos que sean pertinentes y sirvan para encarar las diferentes problemáticas reconocidas y relevantes a nivel local.
- **Periodismo y desarrolladores de contenido.** Por el trabajo que llevan adelante, cuentan con las herramientas y los conocimientos necesarios que les permiten generar información por medio de investigaciones, visualizaciones y otras formas innovadoras que agregan valor a los datos.
- **Emprendedores tecnológicos.** Para ellos, la disponibilidad de datos oficiales siempre es una oportunidad para desarrollar nuevas líneas de negocios, funcionalidades innovadoras y productos o servicios derivados.
- **Academia y comunidad educativa.** Los datos son un recurso educativo importante en todos los niveles de formación. No sólo se pueden convertir en objeto de estudio y material fundamental para el desarrollo de investigaciones aplicadas, sino que además se pueden incorporar al planteamiento de problemas y trabajos prácticos.

02 ¿Cómo abrir los datos?

A la hora de planificar e implementar una política de apertura, te recomendamos crear un plan de implementación con una justificación clara de cómo se realizará. Si bien no hay una única forma de hacerlo -cada organización tiene sus propias dinámicas, estructuras y necesidades- te compartimos algunas consideraciones a tener en cuenta para abrir los datos de un organismo público:

- Seleccionar los datos.
- La conversión de los activos de datos a **formatos abiertos**.
- La estandarización del contenido en base a **entidades interoperables**.
- La generación de **metadatos** para documentar sus activos de datos en un catálogo.

a. Seleccionar los datos

La selección de datos para su apertura requiere del análisis de criterios de valor (para actores internos y externos), costos (para publicarlos y mantenerlos) y riesgos (asociados a su publicación). Algunas preguntas pueden ser de ayuda si querés evaluar el impacto que tendrá la apertura de cada conjunto de datos, no sólo para tu organismo, sino también para la sociedad en su conjunto.

VALOR

Actores. ¿Con qué frecuencia puede ser consumido el recurso? ¿Quiénes son o podrían ser los usuarios internos y externos? ¿Cuántos son? ¿Son los datos de valor a corto o largo plazo para los interesados? ¿Cuánto valor les reporta la interacción con el recurso?

Potenciales usos. ¿Son los datos actualmente utilizados para cumplir funciones del organismo o como soporte de su misión? ¿Los datos incrementan la eficiencia interna del gobierno? ¿Pueden reducir costos? ¿Mejoran la efectividad de los programas de gobierno? ¿Cuál es el potencial para reducir los costos de los actores externos al gobierno? ¿Cuál es el potencial para fomentar la innovación? ¿Cuál es el potencial para crear valor agregado o fomentar el crecimiento económico? ¿Cuál es el potencial para fomentar nuevas formas de colaboración?

COSTOS

Publicación. ¿Se requiere un cambio de formato para compartir los datos? ¿Hay definiciones de los datos y contexto suficiente para una comprensión adecuada? ¿Cuál es el tiempo para su apertura? ¿Se requieren procesos adicionales para compartir los datos? ¿Hay aspectos regulatorios importantes a tener en cuenta para compartir los datos? ¿Cuál es la necesidad de asesoramiento legal para evaluar la apertura? ¿Cuál es el costo global estimado de prepararlos para su apertura?

Operación, actualización y mantenimiento. ¿Qué organizaciones van a comprometer recursos humanos y financieros para compartir los datos? ¿Cómo se identificarán los cambios en los datos que requieran de su actualización? ¿Con cuánta frecuencia se requerirá actualizar los datos? ¿Se requieren nuevos recursos tecnológicos? ¿Cuál es el costo estimado de estos cambios? ¿Cuál es el costo estimado global de mantener los datos? ¿Existen limitaciones u obligaciones internacionales sobre apertura de estos datos? ¿Los datos están protegidos por leyes específicas? ¿Cuáles son las consecuencias potenciales de una mala interpretación de los datos? ¿Es creíble la fuente de los datos?

RIESGOS

Confiabilidad. ¿Pueden los datos inducir a interpretaciones erróneas que disminuyan la confianza en los mismos?

Legales. ¿Los datos pueden comprometer la seguridad pública? ¿Revelan detalles sensibles de infraestructura crítica? ¿Existen limitaciones u obligaciones internacionales sobre apertura de estos datos? ¿Los datos están protegidos por leyes específicas? ¿Los datos pueden plantear un riesgo de seguridad para individuos o grupos de individuos combinados con otros datos existentes?

Sociales y políticos. ¿Los datos liberados pueden tener consecuencias contraproducentes sobre individuos o grupos de individuos? ¿Cuáles son las consecuencias potenciales de una mala interpretación de los datos?

b. Formatos abiertos

Los activos de datos -es decir, los archivos más su metadatos- deben publicarse en formatos que faciliten la reutilización por parte de distintos tipos de usuarios. Esto implica:

- **Elegir formatos de archivos que sean abiertos y procesables por máquinas.** Un archivo CSV (comma-separated values) es un formato específicamente diseñado para publicar datos tabulares y se puede abrir con una multiplicidad de aplicaciones. Por el contrario un PDF (Portable Document Format) no permite procesar datos para su utilización porque es un formato pensado para la producción de documentos (no para la publicación de datos). Un archivo XLSX (Excel) que siga buenas prácticas es un formato mucho mejor, pero está pensado para una variedad de usos más allá de la publicación de datos y no siempre es fácil o seguro utilizarlos para eso.
- **Fragmentar archivos adecuadamente.** Siempre debe haber un archivo que tenga el activo de datos histórico y completo. Sólo algunas veces puede ser conveniente además agregar fragmentos menos pesados del mismo para facilitar su procesamiento a algunos usuarios.
- **Nombrar bien los archivos.** Es importante ser claro y representativo en el nombre de los archivos para que estos puedan ser identificados de forma única en distintos contextos.
- **Nombrar bien las columnas.** Más importante aún que nombrar bien los archivos es hacer lo propio con las columnas. Una vez que un usuario abre una tabla es muy importante que el nombre de la columna abra fácilmente en cualquier aplicación y sea claro y fácil de entender.



Conocé más

Guía para la publicación de datos en formatos abiertos. Recomendaciones para publicar datos en formatos abiertos, estructurar bien una tabla, nombrar archivos y columnas, usar estándares básicos y trabajar con planillas de cálculo.



GUÍA DE BUENAS PRÁCTICAS

Formatos abiertos

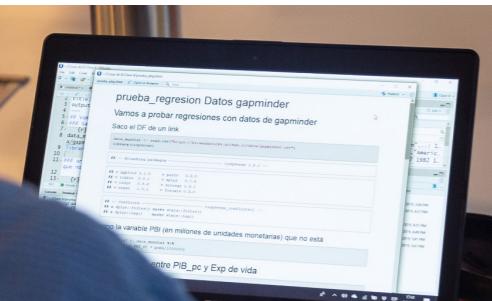
Recomendaciones para la estandarización y publicación de datos



c. Entidades interoperables

Las entidades interoperables son aquellas que se repiten y usan frecuentemente dentro de datasets de temáticas diversas entre sí o de una misma temática (ej: Salud), pero no de otras (como Educación, Economía, Transporte, etc).

La mayoría de los datasets incluyen campos que responden a dónde, quién, cuándo y qué, de modo que sea posible la interoperabilidad, es decir, que los datasets puedan hablar entre sí. Esto no puede suceder cuando dos datasets nombran de forma distinta a una misma entidad interoperable, como cuando se usan distintos sistemas de identificadores o se nombra una misma entidad con o sin mayúsculas, cuando se usan o se dejan de usar artículos y preposiciones, abreviaturas, siglas, tildes, la forma corta o completa de un nombre.



Algunas entidades interoperables importantes que deben estandarizarse son:

- **Unidades territoriales:** provincias, departamentos, municipios, radios y fracciones censales.
- **Direcciones y lugares:** deben escribirse siguiendo una estructura normalizada para maximizar su claridad y las posibilidades de georreferenciación.
- **Personas físicas:** debe usarse el CUIL como identificador único y usar siempre los mismos campos para guardar nombre y apellido.
- **Personas jurídicas:** debe usarse el CUIT como identificador único y usar siempre los mismos campos para el nombre y otros atributos.

Conocé más

Guía para la identificación y uso de entidades interoperables. Recomendaciones y referencias para nombrar entidades en activos de datos, usando sus nombres y códigos oficiales.



d. Metadatos

Si los datos son difíciles de encontrar o contextualizar, de poco sirven. Los metadatos -es decir, los datos sobre los datos- son útiles en la medida en que describen el contenido informativo de un recurso o dataset. Indican aspectos como:

- Clave única de identificación
- Título
- Descripción
- Responsable de publicación
- Temática
- Frecuencia de actualización
- Fecha de publicación
- URL de descarga

A mayor cantidad de metadatos, más sencillo resulta el uso de datos. En la Argentina, existe un esquema de metadatos definido para el uso de todos los organismos de la Administración Pública Nacional, el Perfil Nacional de Metadatos para Datos Abiertos. Busca homogeneizar la forma de generar catálogos de datos para facilitar la búsqueda e indexación de activos de datos en un punto de acceso central: datos.gob.ar.

En el caso de la Administración Pública Nacional, todos sus organismos públicos deben generar o tener acceso a un catálogo donde documentar sus activos de datos.

Conocé más

Perfil de Aplicación Nacional de Metadatos para Datos Abiertos. Especificación para documentar activos de datos digitales de organismos de la Administración Pública Nacional, con el fin de facilitar la búsqueda y acceso mediante la implementación de un catálogo central normalizado, que permita la integración efectiva en sistemas, aplicaciones y flujos de trabajo con datos.

<http://bit.ly/perfilmetadatos>

e. Un portal de acceso

Para facilitar el acceso y navegación a los activos de datos por parte de nuestros usuarios, es importante contar con una plataforma web que los agrupe y facilite su búsqueda y descarga. Independientemente de la plataforma que utilices, es importante garantizar el proceso de documentación de los conjuntos de datos publicados.

ANDINO

Portal Andino es la plataforma de código abierto desarrollada por el equipo de Datos Abiertos de la Administración Pública Nacional de Argentina. Tiene el fin de facilitar el proceso de apertura de datos en los diferentes organismos públicos que quieran encarar el desafío, o estén obligados por alguna norma específica. Es la tecnología detrás del Portal de Datos del Poder Ejecutivo, así como el de varios de sus Ministerios, Secretarías y otros organismos públicos de orden provincial y otros poderes.

Andino es una versión de CKAN, una plataforma de código abierto de OKFN que cuenta con una nutrida comunidad de práctica y ya fue implementado en numerosos gobiernos de distintos países del mundo.

DATOS.GOB.AR

El Portal Nacional de Datos Abiertos (datos.gob.ar) es una plataforma web que funciona como el principal punto de acceso a los datos abiertos por la Administración Pública Nacional. Reúne en un único catálogo todos los activos de datos documentados por la Administración Pública Nacional considerando estándares de calidad y relevancia.

Datos.gob.ar está implementado con Portal Andino.



Andino
Desarrollado con CKAN



04 ¿Cómo crear un catálogo de datos abiertos?

Los catálogos de datos abiertos son listas de datasets -activos con datos y sus metadatos-, que te describen qué son y cómo podés usarlos. Es importante generarlos porque les permiten a los usuarios encontrar y entender los datos publicados por organismos del Estado. Además de ser una herramienta fundamental para promover la trans-

parencia de la gestión, es el medio que habilita el encuentro entre los datos y todos los actores que les agregan valor, lo que permite que surjan nuevos y/o mejores productos y servicios.

Enterate cuáles son los pasos a seguir y dónde podés encontrar más información si querés armar el catálogo de tu organismo.

a. Organismos de la Administración Pública Nacional

Para conformar un catálogo te recomendamos seguir una serie de pasos para abrir los datos y darlos de alta en el Portal Nacional de Datos Abiertos (datos.gob.ar), la plataforma para buscar y acceder fácilmente a los datos que publican los organismos de la Administración Pública Nacional. Esto supone:

- **Elegir una implementación para el catálogo.** Para armar un catálogo, no necesitás tener un portal web, podes crearlo de varias maneras: instalando un Portal Andino o subiendo un archivo de Excel o un JSON a una URL.
- **Seleccionar los activos de datos que se desea abrir.** Tené en cuenta los datasets recomendados por el marco normativo nacional y los índices internacionales. Pero las necesidades específicas de cada organismo son las que guiarán las prioridades teniendo en cuenta cuál será el uso que se dará a los datos publicados. Te recomendamos identificar cuáles son o podrían ser más demandados. Seguí los criterios de valor, costos y riesgos que te compartimos en la sección anterior.
- **Mejorar la calidad de datos y metadatos.** Seguí buenas prácticas para compartir tus datos en formatos abiertos, incorporá rutinas de transformación y limpieza de datos, estandarizalos y enriquecelos para que puedan cruzarse fácilmente con otros.
- **Integrar el catálogo a datos.gob.ar.** Asegurate de que tu catálogo cumple con el Perfil Nacional de Metadatos para Datos Abiertos y esribinos a datos@modernizacion.gob.ar con la URL de tu catálogo para que te sumemos a datos.gob.ar. El equipo de Datos Argentina analiza cada dataset nuevo que publicás y te ayuda a mejorar la calidad de sus datos y metadatos.
- **Publicar datos como 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.

Conocé más

La apertura de datos en la Administración Pública Nacional. Guía para la implementación.

Te compartimos los recursos necesarios para publicar datos en los organismos de la APN. Aprendé cómo mejorar la calidad, identificar activos de datos y publicarlos como un servicio.



b. Provincias y municipios de Argentina

Para facilitar la implementación de la política de apertura de datos en gobiernos provinciales o locales, desde el equipo de Datos Argentina elaboramos algunas recomendaciones:

Conformar un equipo. Estas iniciativas son llevadas a cabo por equipos reducidos y dinámicos, con la capacidad de ajustarse rápidamente a las necesidades del entorno. Identificamos algunos roles y perfiles claves, tales como el líder político, el coordinador, el perfil técnico y uno analista a tener en cuenta para formar los equipos a nivel municipal o provincial.

Identificar datos prioritarios para la apertura. Tené en cuenta los datasets recomendados por el marco normativo nacional y los índices internacionales. Pero las necesidades específicas de cada gobierno provincial o local, guiarán las prioridades teniendo en cuenta cuál será el uso que se dará a los datos publicados. Seguí los criterios de valor, costos y riesgos que te compartimos en la sección anterior.

Realizar una publicación de datos basada en buenas prácticas. Proponete incluir datos que cumplan con los estándares locales -y en lo posible nacionales e internacionales- que ya existen para datos abiertos. Considerá, para empezar, expresar con claridad los principios de la Carta Internacional de Datos Abiertos (Open Data Charter, 2015) en tu política. Además asegurate de elegir formatos abiertos, de estandarizar el contenido para que sea interoperable y que los metadatos te permitan documentar los activos.

Encontrar y conectar a la comunidad que usa datos abiertos. Es necesario crear una comunidad con los actores que los reutilizan, promover espacios de colaboración y generar distintas instancias en las que las personas que poseen estos conocimientos se acerquen a los datos.

Conocé más

La apertura de datos en gobiernos provinciales y locales. Guía para la implementación. Si formás parte del equipo de trabajo de una provincia o gobierno local, te compartimos las bases y los conceptos claves para que puedas diseñar e implementar políticas de datos abiertos.





Contacto

Dirección Nacional de Datos e Información Pública
Subsecretaría de Innovación Pública y Gobierno Abierto
Secretaría de Modernización

www.argentina.gob.ar/datospublicos
datos@modernizacion.gob.ar

medium.com/@datosgobar
github.com/datosgobar

