

KIT de DATOS ABIERTOS



Ministerio de Modernización
Presidencia de la Nación

ÍNDICE

1

Introducción

2

Datos abiertos

- a. ¿Qué son los datos abiertos?
- b. Características de los datos abiertos.
- c. Diferencia entre datos e información.
- d. Diferencia entre datos públicos y datos abiertos.

3

¿Por qué abrir los datos públicos y convertirlos en datos abiertos?

- a. Las 8 características de los datos abiertos.
- b. Niveles de apertura de los datos: las 5 estrellas.

4

Recomendaciones básicas y herramientas para iniciar la apertura de datos

- a. ¿Cómo identificar los datos?
- b. ¿Cómo abrir los datos?
 - Seleccionar los datos.
 - Aplicar una licencia abierta.
 - ¡Abrir los datos!
 - ¡Promover el uso!

5

Comunidades y aportes

6

El idioma de los datos

Introducción

01

En el último tiempo las Administraciones Públicas en todo el mundo enfrentan nuevos desafíos producto de la introducción de Tecnologías de Información y Comunicación (TIC). Se presentan diferentes formas de relacionamiento y demandas, cada vez más específicas, a las que los Gobiernos deben adaptarse y generar un profundo cambio interno y en su relación con la comunidad.

El Plan de Modernización del Estado impulsado por el Gobierno Nacional tiene como objetivo avanzar hacia una Administración Pública al servicio del ciudadano, promoviendo un Estado sólido, moderno y eficiente, con equipos técnicos profesionalizados y capacitados para brindar soluciones y resultados dentro de un marco de ética, transparencia y rendición de cuentas.

La Subsecretaría de Innovación Pública y Gobierno Abierto del Ministerio de Modernización de la Nación desarrolló una serie de Kits de Herramientas que les permiten a los empleados y funcionarios del Estado hacer frente a los desafíos de manera práctica, a través de la aplicación de los principios de apertura e innovación pública.

En este Kit desarrollamos conceptos y herramientas vinculadas a los datos abiertos, sus características y beneficios. Además, compartimos el camino ideal para empezar a abrir los datos en un organismo. También analizamos la importancia de la construcción de una comunidad que pueda aprovechar la información y generar valor con ella. ¡Bienvenidos al mundo de los datos abiertos!

02 Datos abiertos

a. ¿Qué son los datos abiertos?

Aún sin darnos cuenta, **todo el tiempo nos rodeamos de datos**. Desde las planillas que abrimos en la oficina del trabajo, hasta las investigaciones que llevan adelante los científicos que buscan curas a enfermedades.

Así como nosotros lo hacemos, en el ejercicio de sus funciones la **Administración Pública produce grandes cantidades de datos**. Éstos son de todas las personas y no pertenecen a ningún Gobierno. Su valor es enorme: **los distintos sectores de la sociedad pueden tomar decisiones de muchísima importancia con ellos**. Por eso, y siempre que no contengan información personal sobre individuos específicos que pongan en riesgo su privacidad, es fundamental abrirlos.

Los datos abiertos son aquellos a los que cualquier persona puede acceder, usar y compartir libremente.

Sólo deben atribuirse y compartirse con la misma licencia con la que fueron publicados.

**LOS DATOS ABIERTOS SON AQUELLOS
A LOS QUE CUALQUIER PERSONA PUEDE
ACCEDER, USAR Y COMPARTIR LIBREMENTE.**

“Los datos por sí mismos no valen nada si no hay un problema para resolver y una comunidad interesada en resolverlo.”

Beth Noveck, The GovLab

Datos abiertos 02

b. Características de los datos abiertos

Tienen que cumplir con tres atributos básicos:

- **Disponibilidad y acceso:** estar disponibles en su totalidad y sin costo de reproducción, preferiblemente tienen que poder descargarse de manera gratuita en Internet.
- **Reutilización y redistribución:** permitir realizar productos derivados, combinarlos con otras fuentes de información y distribuirlos de forma gratuita.
- **Participación universal:** todas las personas deben poder utilizar, reutilizar y redistribuir la información. No debe haber discriminación alguna en términos de esfuerzo, personas o grupos. Restricciones “no comerciales” que prevendrían el uso comercial de los datos, o restricciones de uso para ciertos propósitos (por ejemplo, sólo para educación) no son permitidos.

Un DATO es un elemento descontextualizado.

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

La INFORMACIÓN es un dato procesado, con valor agregado.

Son, por ejemplo, las estadísticas creadas sobre la cantidad de usuarios que usan una línea de subte.

c. Diferencia entre dato e información

DATO

El dato es un elemento descontextualizado que puede dar origen a la generación de información. Son 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. Los datos son la verdadera materia prima para generar información con valor agregado.

INFORMACIÓN

Es un dato dotado de relevancia y utilidad. Para obtener información es necesario aplicar algún tipo de transformación o procesamiento sobre los datos. Convertir datos en información es un proceso que genera valor, aporta relevancia, genera algo nuevo que se desconocía.

d. Diferencia entre dato público y dato abierto

Dato público: es cualquier dato generado en el ámbito gubernamental, o que se encuentra bajo su guarda, y que no tiene su acceso restringido por ninguna legislación específica.

Dato abierto: es el dato público disponibilizado en un medio digital, bajo una licencia abierta y usando para ello un formato estándar abierto.

¿Por qué abrir datos públicos y convertirlos en datos abiertos?

03

El sector público, los privados, las organizaciones de la sociedad civil, las universidades y/o cualquier ciudadano, puedan reutilizar los datos, agregarles valor, generar información y crear nuevos o mejores servicios y aplicaciones.

BENEFICIOS DE LA APERTURA DE DATOS

Mejora el proceso de toma de decisiones

Permiten tomar decisiones basadas en evidencia. La identificación, ordenamiento y gestión de la información en la Administración Pública mejoran la interoperabilidad y la comunicación interna.

Hace posible la transparencia y la rendición de cuentas

Abrir información transparente la gestión y 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 que están 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, a partir de poder 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, la reducción de costos y la posibilidad de abrir nuevas fuentes de trabajo gracias a los descubrimientos posibilitados por los datos.

“Queremos que la información sea usada por la sociedad civil y así constituir mecanismos colaborativos que faciliten desarrollos estadísticos y aplicaciones”.

Andrés Ibarra, Ministro de Modernización.

a. Los 8 principios de los datos abiertos

La iniciativa *Open Government Data*, definió **ocho** características que especifican condiciones a tener en cuenta para que los datos “públicos” sean, además, efectivamente “abiertos”:

- 1 **Completos:** están disponibles y no tienen restricciones para su apertura.
- 2 **Primarios:** son recolectados en la fuente de origen, con el nivel de especificidad más alto posible, no en forma agregada ni modificada.
- 3 **Oportunos:** están disponibles tan rápido como sea necesario para preservar su valor.
- 4 **Accesibles:** están disponibles para el rango más amplio de usuarios y propósitos.
- 5 **Procesables por máquinas:** están estructurados de forma tal que permitan un procesamiento automático.
- 6 **No discriminatorios:** están disponibles para cualquier ciudadano, sin requerir un registro.
- 7 **No propietarios:** están disponibles en un formato sobre el cual ninguna empresa de software tiene un control exclusivo.
- 8 **Licencias libres:** los datos abiertos deben tener una licencia que especifique que se trata de datos abiertos y que defina bajo qué condiciones estos datos pueden ser reutilizados. La licencia podría, además, detallar que las personas que usen los datos deben atribuirlos y que, en caso de que combinen los datos con otros datos, deben compartir esta información también como datos abiertos. Más adelante, en *Aplicar una licencia abierta*, veremos cuáles son las mejores opciones.

¿Por qué abrir datos públicos y convertirlos en datos abiertos? 03

b. Niveles de apertura de los datos: **las 5 estrellas**

Existen diferentes escalas y parámetros para definir el nivel o grado de apertura de los datos. Tim Berners-Lee, el inventor de la Web e iniciador de los Datos Enlazados (Linked Data), sugirió un esquema de clasificación a partir de 5 estrellas, cada una de las cuales representa un nivel o escalón de desarrollo en la apertura de los datos, del más simple al más complejo.



Primera estrella: ¡mucho por mejorar!

Cuando el formato en el que publicamos los datos no permite su reutilización, por ejemplo un pdf o una imagen escaneada, nos encontramos en presencia del primer nivel de apertura de datos.



Segunda estrella: ¡todavía cuesta leer los datos!

Ya en el segundo nivel de dos estrellas, los usuarios pueden operar sobre los datos y agregar valor. Sin embargo, aún estamos en presencia de datos que requieren de un software propietario para su lectura.



Tercera estrella: el mínimo que buscamos

La tercera estrella significa que logramos publicar los datos en formatos estructurados, no propietarios, que pueden ser utilizados sin necesidad de pagar por un software. ¡Este es el estadio mínimo que necesitamos al abrir los datos!



Cuarta y quinta estrella: hacia la excelencia

La cuarta y quinta estrella son niveles más complejos de apertura, que requieren de quien los publica, mayor nivel de tecnología y desarrollo de infraestructura de información para lograr calidad y homogeneidad en la publicación. Algunos requisitos son publicar los datos en direcciones web únicas, y vincularlos entre sí para hacer inteligencia y agregado de valor.



Recomendaciones básicas y herramientas para iniciar la apertura de datos.

04

a. Identificar los datos

Primero lo primero. Se trata de **identificar los datos que generan, poseen y administran las distintas dependencias públicas.**

b. ¿Cómo abrir los datos?

Una vez identificados los datos, hay que avanzar en el proceso de apertura. La fundación *Open Knowledge International* propone la siguiente metodología a partir de cuatro pasos:

A Seleccionar los datos

B Aplicar una licencia abierta

C Abrir los datos

D ¡Promover el uso!

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).**

Estos criterios nos ayudan a evaluar el impacto que tendrá la apertura de cada conjunto de datos, no sólo para el organismo que la realiza, sino también para la sociedad en su conjunto.

Las siguientes preguntas nos pueden ayudar con esta evaluación:

Criterio	Subcriterio	Pregunta
Valor	Actores	¿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?
		¿Con qué frecuencia puede ser consumido el recurso?
		¿Cuáles son las limitaciones de los de los interesados para el análisis y uso de esos datos?
	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?
		¿Los datos mejoran la efectividad de los programas de gobierno?
		¿Cuál es el potencial de los datos para reducir los costos de los actores externos al gobierno?
		¿Cuál es el potencial de los datos para reducir los costos de los actores externos al gobierno?
		¿Cuál es el potencial de los datos para fomentar la creación de oportunidades de negocio?
		¿Cuál es el potencial de los datos para fomentar la innovación (Ej.: habilitando el desarrollo de nuevas herramientas)?
		¿Cuál es el potencial de los datos de crear valor agregado o fomentar el crecimiento económico?
		¿Cuál es el potencial de los datos para fomentar nuevas formas de colaboración?

criterio	Subcriterio	Pregunta
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 estimado de preparar los datos para su apertura?
		¿Se requieren procesos adicionales para compartir los datos?
		¿Hay aspectos regulatorios importantes a tener en cuenta para compartir los datos (Ej.: privacidad, seguridad, accesibilidad, obligación legal de publicar, etc.)? ¿Cuál es la necesidad de asesoramiento legal para evaluar la apertura de datos?
		¿Cuál es el costo global estimado de preparar los datos 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?
		¿Qué cambios de sistemas deberán implementarse? ¿Se requieren nuevos recursos tecnológicos? ¿Cuál es el costo estimado de estos cambios?
		¿Cuál es el costo estimado global de mantener los datos?
		¿Puede la integridad de los datos se cuestionada? ¿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? (Discriminación laboral por antecedentes de salud, económicos, panales; Persecución cultural, política o religiosa)
		¿Cuáles son las consecuencias potenciales de una mala interpretación de los datos?

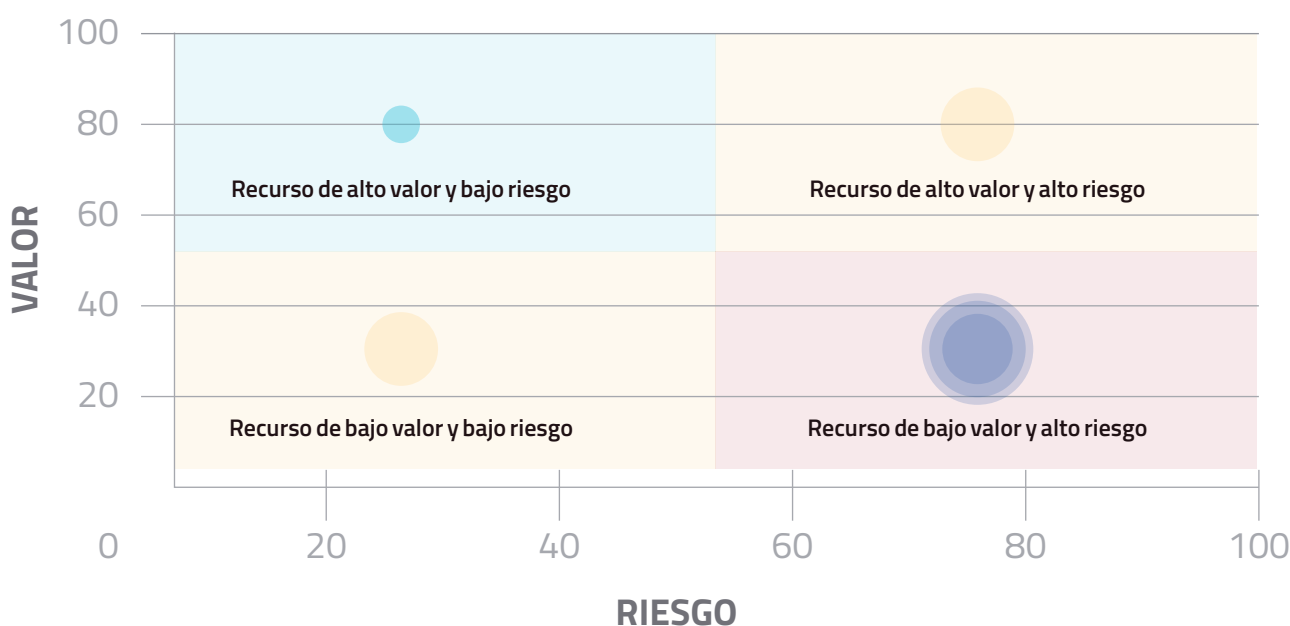
Fuente: elaboración sobre la base del "Open Data Priorization Toolkit" de CIO.gov

¿Cómo ponderar los criterios de la evaluación?

Para priorizar la apertura de datos hay que ponderar los criterios de evaluación (valor, costos y riesgos) y sus subcriterios: (a) el valor que pueden generar para actores internos y externos, (b) los costos para publicarlos y mantenerlos, y (c) los riesgos asociados a su publicación.

Matriz de priorización

Este gráfico se utiliza para ilustrar la relación entre los 3 criterios evaluados en la "Tabla de evaluación de apertura". En el **eje X** se evalúan los riesgos de publicar datos y en el **eje Y** su valor. El tamaño del punto representa los costos de hacerlo.



Si, tras ponderar un conjunto de datos, encontramos que queda posicionado en la matriz en el cuadrante superior-izquierdo (es decir, **con alto valor y bajo riesgo**), debería ser el conjunto de datos elegido para iniciar la apertura, porque el costo de apertura es pequeño.

Por el contrario, si un conjunto de datos resulta posicionado en la matriz en el cuadrante inferior-derecho (**bajo valor y alto riesgo**), entonces ese debería ser el

último conjunto de datos que vayamos a abrir, ya que el costo de apertura es muy grande.

Cada organismo puede establecer sus propios criterios de evaluación y mecanismos de ponderación. Lo importante es iniciar el proceso de apertura desde lo más sencillo a lo más complejo, es decir comenzar a abrir los datos que no requieran altos costos y niveles de complejidad, y que a su vez, aporten valor a la comunidad.

B. Aplicar una licencia abierta

Los datos abiertos necesitan estar abiertos tanto técnica, como legalmente. Para hacerlo, tenemos que aplicarles una licencia.

Los datos en poder del sector público no deben estar sujetos a restricciones de patentes.

Más importante aún es asegurarse de que estén en formatos legibles que permitan una mayor reutilización. Por ejemplo, documentos en PDF (Portable Document Format) limitan la extracción de los datos y su procesamiento.

¿Qué licencias podemos usar?

La fundación *Open Knowledge International* desarrolló las primeras licencias específicas para datos abiertos, las licencias Open Data Commons, según se describen:

- **Open Data Commons Public Domain Dedication and License o PDDL**, que permite difundir, reutilizar o adaptar los datos sin restricción alguna.
- **Open Data Commons Attribution License**, al igual que la licencia de **Creative Commons Reconocimiento**, se exige la referencia a la autoría o fuente de los datos para la reutilización de la información.
- **Open Data Commons Open Database License (ODbL)** que permite la reutilización de los datos siempre que se reconozca la autoría de la información original; se mantenga la misma licencia en las obras derivadas las cuales puede restringir su uso si, además, se distribuye una versión sin dichas restricciones³.
- **Creative Commons Atribución 4.0 Internacional** esta licencia posibilita no sólo copiar y redistribuir el material en cualquier medio o formato, sino también adaptar y transformarlo para cualquier propósito, incluso comercial.

Una breve guía para utilizar una licencia para datos abiertos puede ser encontrada en el sitio Open Data Commons: <http://opendatacommons.org/guide/>

³Portal de Datos Abiertos de España, disponible en <http://datos.gob.es/> consultado el 29 de julio de 2016.

C. ¡Abrir los datos!

Una vez que elegimos qué datos podemos abrir y con qué licencia, llega el mejor momento: compartirlos. Para eso, **necesitamos un catálogo**. Ellos son la herramienta principal para ofrecer los datos al mundo. Consisten en uno o varios portales que facilitan el ordenamiento y acceso a los distintos conjuntos de datos. Los catálogos de datos están formados por distintos elementos como búsqueda, metadatos, información sobre licencias y acceso a los conjuntos de datos.

Podemos encontrar diferentes frameworks de software que nos permitirán construir nuestros catálogos, como CKAN (utilizado por Datos Abiertos Argentina www.datos.gob.ar), DKAN (utilizado por el Ministerio de Educación de Perú datos.minedu.gob.pe), Socrata (utilizado por la Casa Blanca open.whitehouse.gov/browse) y Junar.com (usado por el municipio de Bahía Blanca gabierto.bahia-blanca.gob.ar/datos-abiertos/)

D. ¡Promover el uso!

Los datos abiertos no son nada sin los usuarios. Es fundamental que los datos puedan ser encontrados con facilidad y que exista una comunidad que pueda re-utilizarlos para generar valor a partir de ellos.

Cuando abrimos los datos, necesitamos empezar a **construir comunidades que usen la información**. Algunas, como las de desarrolladores o visualizadores, **son esenciales en la primera etapa de apertura de datos, ya que a través de ellos se pueden crear grandes herramientas que, a su vez, lleguen al público en general.**

Existen organizaciones, comunidades y especialistas en el uso y explotación de los datos abiertos, como por ejemplo los investigadores, periodistas de datos y comunidades de desarrolladores informáticos.

Crear comunidades que usen los datos es clave. Algunas, como las de desarrolladores o visualizadores, son esenciales en la primera etapa de apertura de datos, ya que a través de ellos se pueden crear grandes herramientas que, a su vez, lleguen al público en general.

Comunidades y aportes

Desarrolladores

La creación de aplicaciones a través de datos abiertos es casi ilimitada. ¿Quién no usó una aplicación para saber qué medio de transporte nos lleva de casa al trabajo, comparando tiempos y costos?

Visualizadores

Una buena visualización que ponga de relieve a través de gráficos el uso del presupuesto de un gobierno es una forma amigable de acercar los datos públicos y abiertos a todos los ciudadanos.

Privados

Muchísimas son las empresas que usan datos abiertos para ofrecer sus servicios. Por ejemplo, aquellas que permiten a sus clientes encontrar una casa en venta o en alquiler teniendo en cuenta la distancia hasta transportes o escuelas públicas.

Periodistas de datos

Investigaciones sobre declaraciones juradas y chequeos de discursos, son dos de los principales ejemplos de periodismo de datos. Es una práctica fundamental para hacer más transparente el funcionamiento de la Administración Pública.

El idioma de los datos

Estos son algunos conceptos frecuentes en el mundo de los datos abiertos. Te invitamos a familiarizarte con ellos.

○ Activo de información.

Es cualquier recurso de información o datos con valor para el desarrollo de las funciones de una organización, que puede ser comprendido y tratado como una única unidad a efectos de gestión, uso, protección e intercambio. Aunque puede designar piezas aisladas de información (una imagen incluida en un documento, un registro de una base de datos), suele emplearse para identificar y tratar conjuntos de información o datos, como agrupaciones documentales, bases de datos, sitios web, colecciones de metadatos. Por ejemplo, una base de datos.

○ Catálogo de datos

Es el sistema de información que recopila y organiza metadatos descriptivos sobre los distintos conjuntos de datos abiertos por una organización, proporcionando un acceso único a los mismos. Por ejemplo, un portal de datos.

○ Conjunto de datos o datasets

Son la pieza principal en todo el catálogo de datos. Éstos agrupan uno o más recursos de datos y son el eje sobre el que se documenta la metadata de la información.

○ Recursos

Son la pieza de información más pequeña del catálogo y son los verdaderos activos de datos del portal.

Autoridades

Presidente de la Nación

Mauricio Macri

Director Nacional de Datos e Información Pública

Gonzalo Iglesias

Ministro de Modernización

Andrés Ibarra

Directora de Datos Públicos

Natalia Sampietro

Secretaria de Gestión e

Innovación Pública

María Inés Baqué

Subsecretario de Innovación Pública y

Gobierno Abierto

Rudi Borrman

Equipo de trabajo

Dirección

Gonzalo Iglesias

Desarrollo y redacción

Natalia Sampietro, Carolina Raspo, Pilar Mayora, Marisol Parnofiello, María Sol Tischik, Vanesa Bentrón, Claudia Román.

Diseño

Ignacio Leguizamón, Ornella Falcone y Manuel Mercúri.

Impresión

Emilio Etchart, Pablo Conde y Ricardo Gamarra.



Este trabajo está disponible bajo la licencia Creative Commons Atribución-No Comercial-Licenciamiento Recíproco 4.0 Internacional.



Ministerio de Modernización Presidencia de la Nación



www.twitter.com/GobAbierto_AR



www.facebook.com/GobAbiertoAR



www.argentina.gob.ar/modernizacion