

**HANDBOOK**

---

# Estado abierto a través de Datos abiertos

---

REPÚBLICA ORIENTAL DEL URUGUAY  
PROVINCIA DE MENDOZA, ARGENTINA

DRA. JAVIERA ATENAS  
EDICIÓN



**INDES**

**<>agesic**  
DESARROLLANDO  
EL URUGUAY DIGITAL





Esta obra está bajo una Licencia Creative Commons  
Atribución-NoComercial 4.0 Internacional (CC BY-NC 4.0)

Usted es libre de compartir (copiar y redistribuir el material en cualquier medio o formato) y adaptar (remezclar, transformar y construir a partir del material); bajo los siguientes términos: Atribución (Usted debe dar crédito de manera adecuada, brindar un enlace a la licencia, e indicar si se han realizado cambios. Puede hacerlo en cualquier forma razonable, pero no de forma tal que sugiera que usted o su uso tienen el apoyo de la licenciante.) NoComercial (Usted no puede hacer uso del material con propósitos comerciales).

**Edición:** Javiera Atenas.

**Contenidos:** Juan Belbis, Silvia de Rosa, Juan Pane, Carla Bonina, Juan Casanueva, Camila Salazar, Daniel Villatoro, Sergio Araiza y Javiera Atenas.

**Revisión de estilo:** Daniel Villatoro y Camila Salazar

**Diseño y publicación:** Daniel Villatoro

**DOI:** <http://doi.org/10.5281/zenodo.3243389>

**Citar como:** Atenas, J. (Ed.). (2019). Estado abierto a través de datos abiertos. Montevideo, Uruguay: Iniciativa Latinoamericana por los Datos Abiertos y Banco Interamericano de Desarrollo. <http://doi.org/10.5281/zenodo.3243389>



# Contenidos

## MÓDULO 1

### Gobierno Abierto y datos abiertos

05

Docente: Juan Belbis

	Resumen introductorio	06
Unidad 1	Gobierno Abierto y Estado Abierto	07
Unidad 2	Ética, privacidad y apertura de datos	11
Unidad 3	Licenciamiento de la información	16
Unidad 4	Valor de publicación y uso de datos abiertos	19
Unidad 5	Acceso a la información pública y datos abiertos	24
	Lecturas	30

## MÓDULO 2

### Publicando Datos Abiertos desde el sector público

31

Docente: Dr. Juan Pane

	Resumen introductorio	32
Unidad 1	Desarrollo de una estrategia de liberación de DatosAbiertos	34
Unidad 2	Indicadores internacionales de datos abiertos y esquema de publicación	38
Unidad 3	Arquitecturas y alternativas de plataformas para liberar datos	43
Unidad 4	Estándares para liberación de datos e instrumentos de medición	48
Unidad 5	Estándares de datos de áreas específicas	50
	Lecturas	55

## MÓDULO 3

### ¿Para qué liberar datos? Entendiendo la demanda

56

Docente: Dra. Carla Bonina

	Resumen introductorio	57
Unidad 1	El valor de la tecnología cívica y los datos abiertos	59
Unidad 2	Quiénes piden datos abiertos, quiénes los usan y para qué	63
Unidad 3	Datos Abiertos para la mejora de servicios públicos	67
Unidad 4	Datos abiertos para la economía: innovación, emprendedores y datos	71
Unidad 5	Datos abiertos y políticas públicas	76
	Lecturas	79

## MÓDULO 4

### Desarrollando habilidades prácticas para trabajar con datos

80

Docente: Juan Casanueva

	Resumen introductorio	81
Unidad 1	Trabajando con herramientas de visualización	84
Unidad 2	Trabajando con geo-referenciación	91
Unidad 3	Calidad de los datos	93
Unidad 4	Una mirada básica a estadística y análisis de datos	99
Unidad 5	Introducción al periodismo de datos	107

## MÓDULO 5

### Usos de datos en mi organización

112

Docente: Dra Javiera Atenas

	Resumen introductorio	113
Unidad 1	Priorización en la apertura de datos	114
Unidad 2	Introducción a los criterios para la apertura de datos	118
Unidad 3	Introducción al desarrollo de planes de publicación de Datos Abiertos	122
Unidad 4	Introducción al diseño centrado en las necesidades de los usuarios.	125
Unidad 5	Formulación de proyectos de datos abiertos a nivel organizacional	128
	Lecturas	132

### Agradecimientos

133

## MÓDULO 1

---

# Gobierno abierto y datos abiertos

---

Docente: Juan Belbis  
Contribuciones: Silvia Da Rosa  
Javiera Atenas  
Juan Pane

TUTORIAS



Manuel Podetti  
Silvia Da Rosa



Romina Colman  
Ignacio Alfaro

# Resumen introdutorio

Este módulo introducirá a los participantes en temáticas relacionadas con el gobierno abierto y su conexión con las políticas de datos abiertos. Los objetivos de este módulo se centran en facilitar a los participantes la comprensión de conceptos y su relación con casos de estudio. Esta teoría funda las bases de los módulos siguientes, lo cual permitirá que los estudiantes puedan progresar efectivamente en el curso. En particular, se introducirán los siguientes temas:

- Gobierno Abierto y Estado Abierto
- Ética, privacidad y apertura de datos
- Licenciamiento de la información e introducción a las licencias para uso de datos abiertos
- Valor de publicación y uso de datos abiertos
- Acceso a la información pública y datos abiertos

## OBJETIVOS DE APRENDIZAJE

1. Distinguir los principales elementos detrás del concepto de gobierno abierto.
2. Comprender el concepto de acceso a la información pública.
3. Explorar los debates éticos sobre el uso de datos por parte del sector público.
4. Identificar el valor de la apertura y uso de los datos.
5. Comprender la relación entre el acceso a la información pública y la agenda de datos abiertos gubernamentales.

## GLOSARIO DE TÉRMINOS

**Acceso a la información Pública:** se entiende como la capacidad de solicitar y obtener información producida y resguardada por los gobiernos, con sujeción a las excepciones definidas en la regulación de la materia.

**Dato:** Componente mínimo que conforma, luego de su contextualización y análisis, lo que conocemos como información. Un dato puede ser una letra, un número o cualquier símbolo que representa una cantidad, una medida, una palabra o una descripción.

**Datos Abiertos Gubernamentales:** Datos producidos y almacenados por los gobiernos que se encuentran disponibles en formatos digitales abiertos que permiten su reutilización.

**Gobierno Abierto:** un conjunto de políticas públicas orientadas a la mejora de prácticas gubernamentales que tengan como principios la transparencia, la participación y la colaboración

**Información:** conjunto organizado de datos procesados que transmiten un mensaje/idea.

**Formato abierto:** son formatos de archivo no propietarios, cuya especificación está documentada públicamente o es de libre conocimiento e implementación. Los formatos abiertos están libres de patentes y de cualquier otra restricción legal o económica para su utilización.

**Licencia abierta:** Los datos abiertos deben tener una licencia que especifique que se trata de datos abiertos, y defina bajo qué condiciones pueden ser reutilizados

## Introducción

¿Qué es el gobierno abierto? ¿Qué distingue a un gobierno abierto de uno cerrado? ¿Cuándo surge este concepto? Pretendemos contestar estas preguntas en este primer capítulo del primer módulo. Intentaremos compartir conceptos y datos para construir entre todos una visión unificada sobre las políticas de apertura gubernamental.

Comprendemos al gobierno abierto como un conjunto de políticas públicas orientadas a la mejora de prácticas gubernamentales que tengan como principios la transparencia, la participación y la colaboración. Los orígenes de estas prácticas en las administraciones públicas contemporáneas, remiten al primer memorándum de la gestión Obama, publicado el 21 de enero de 2009. En ese memorándum se plantean los tres ejes de la transparencia, la participación y la colaboración como los centrales de su gestión:

“Mi administración está comprometida a crear niveles de apertura como nunca antes en el Gobierno. Trabajaremos para asegurar la confianza pública y establecer un sistema de transparencia, participación pública y colaboración. La apertura fortalecerá nuestra democracia y promoverá la eficiencia y la eficacia en el Gobierno”. (Obama, 2009)

Estos principios son muy amplios y poco claros en su aplicación práctica, con lo cual deberíamos hacer una primera interpretación sobre qué implican cada uno de ellos, desde la gestión pública y la sociedad civil.

Pilar conceptual	Aplicación en la gestión	Rol social y cívico
Transparencia	Acceso a la información pública y apertura de datos públicos.	Ejercer control social (rendición de cuentas) y reutilización de información pública para promover innovación y desarrollo económico.
Participación	Plataformas y mecanismos de diseño e implementación de políticas públicas integrando actores sociales.	Incidencia directa de la ciudadanía en la toma de decisiones gubernamentales.
Colaboración	Generación de espacios de interacción e innovación dentro del estado, y entre el estado, la sociedad civil y actores privados	Co-diseño de soluciones y herramientas que generen valor público

El Gobierno Abierto reúne estos principios, e incorpora prácticas innovadoras para concretarlos. La tecnología y el contexto de la sociedad de la información sirven para dar valor a estas políticas públicas y reducir las brechas entre la ciudadanía y los actores públicos.

## **1.1 Transparencia y rendición de cuentas**

La transparencia es la condición para comenzar un proceso de diálogo constructivo entre la ciudadanía y las instituciones gubernamentales. Sin instituciones transparentes la rendición de cuentas es de baja calidad y el escrutinio público se hace imposible. Esta ponderación sobre la importancia de la transparencia en las políticas de apertura, genera muchas veces de forma equivocada la impresión de que apertura es sinónimo de transparencia.

La transparencia es condición necesaria para la implementación de políticas de apertura, pero no suficiente. Más adelante en este módulo veremos las dos dimensiones para implementar la transparencia como política pública: la pasiva a través de la garantía del derecho al acceso a la información pública, y la activa promovida en general como políticas de datos abiertos gubernamentales.

## **1.2 Participación ciudadana**

Apertura implica participación, en términos amplios y concretos. Por ejemplo, presupuestos participativos, procesos de co-creación de políticas públicas, espacios de debates y audiencias públicas, plataformas de iniciativas ciudadanas. Algunos de estos planteamientos se potencian con la incorporación de tecnologías que facilitan la participación remota y reducen las brechas de acceso a las instituciones públicas. Otros tienen un componente analógico tan antiguo como el concepto de la democracia. La participación eleva la calidad de nuestras democracias e incorpora recursos a las discusiones públicas con el fin de mejorar el rendimiento de las instituciones.

## **1.3 Colaboración inter e intragubernamental**

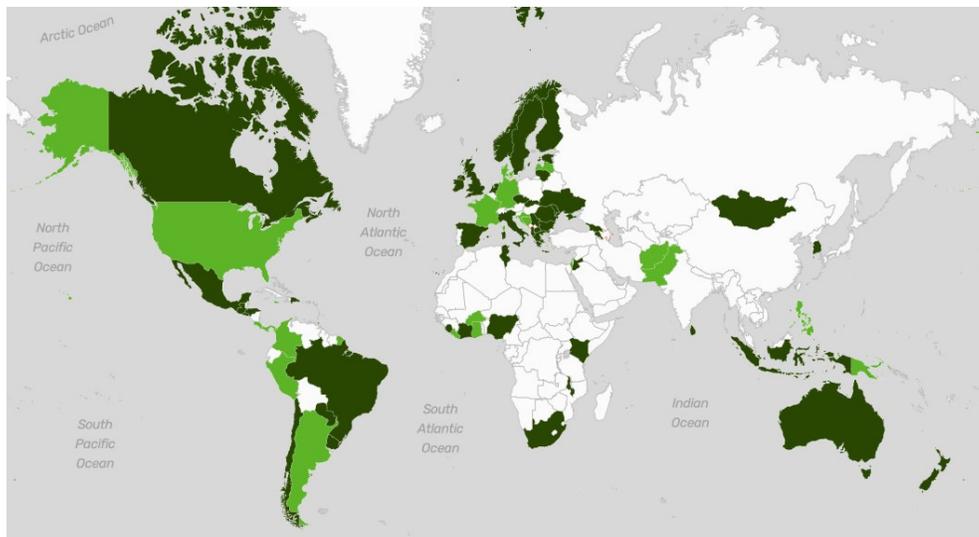
Así como se pretende incorporar al ciudadano de forma activa en los procesos de la política pública, es importante romper la lógica vertical de la administración pública fomentando la colaboración inter e intragubernamental. Los organismos públicos no son estamentos aislados, sino que integran el sistema de la administración pública en diferentes sentidos. Este pilar apunta hacia la mejora de los servicios públicos, la reducción concreta de duplicidad de tareas en las instituciones públicas y una mejora en la eficiencia de la administración pública en general.

## **1.4 La Alianza por el Gobierno Abierto (OGP)**

Las definiciones de gobierno deben considerar el rol de la Alianza para un Gobierno Abierto (AGA), creada en el 2011, cuando Estados Unidos, Reino Unido, Brasil, México, Sudáfrica, Indonesia, Noruega y Filipinas, respaldaron la Declaración de Gobierno Abierto y comenzaron a difundir sus buenas prácticas. Esta declaración se basa en cuatro puntos principales:

- Incrementar la disponibilidad de información sobre las actividades gubernamentales.
- Promover la participación cívica.
- Implementar los más altos estándares de integridad profesional a través de las administraciones.
- Incrementar el acceso a las nuevas tecnologías para la apertura y la rendición de cuentas.

La participación de un país en la AGA implica que debe pasar por un proceso en el cual cada dos años debe presentar un Plan de Acción Nacional (PAN) que reúne un conjunto de compromisos orientados a mejorar sus niveles en los cuatro puntos anteriores. Este proceso está acompañado de una metodología que requiere la participación de la sociedad civil y el sector privado en la definición de estos compromisos, y también en sus procesos de implementación.



Fuente: Open Government Partnership

A abril de 2019, 75 países han participado de la AGA y presentado sus respectivos PAN. Adicionalmente hay un programa piloto orientado en el que hasta el momento participan 15 gobiernos subnacionales incluyendo ciudades y provincias. En total se han presentado más de 2500 compromisos asumidos de forma conjunta entre los gobiernos y la sociedad civil.

### 1.5 Hacia un Estado Abierto

Existe una particularidad propia del idioma castellano sobre la adopción de estos conceptos en América Latina, que está determinada por la traducción del término *government*. En la región entendemos al gobierno como la estructura del Poder Ejecutivo, mientras que en la cultura anglosajona el término hace referencia al Estado en su conjunto, incluyendo también a los poderes legislativo y judicial. Por ello, diferentes autores han hecho hincapié en la necesidad de que estas buenas prácticas en el diseño y ejecución de políticas públicas, se reproduzcan también en el resto de los poderes.

La apertura parlamentaria y la apertura judicial se encuentran en procesos más graduales de adopción, ya que su organización institucional rígida en ocasiones dificulta la toma de decisiones que impliquen cambios culturales importantes. No obstante, existen casos destacados de apertura en estas áreas en la región.

#### Parlamento abierto

Existen casos destacados en la adopción de políticas de apertura en los poderes legislativos. Un ejemplo es el Laboratorio Hacker de la Cámara de Diputados de Brasil, un espacio tanto físico como digital en el cual se promueven diferentes iniciativas de innovación pública aplicadas al contexto legislativo. Como parte del proyecto, se organizan hackatones,

desarrollo de aplicaciones, debates y experimentos que buscan integrar al ciudadano en los procesos legislativos. También es importante el rol de organismos como Parlaméricas (una red de parlamentarios de todo América, que promueve diversas agendas, entre ellas la de apertura parlamentaria) que promueve estas prácticas en la región y disemina los casos de éxito entre sus miembros.

#### Justicia abierta

Los poderes judiciales tienen ciertas características propias que suelen dificultar la apertura de sus procesos e información. Sin embargo existen casos de iniciativas exitosas.

En Argentina existe el programa Justicia Abierta en el marco del Ministerio de Justicia de la Nación, que realizó el acuerdo junto con los poderes judiciales subnacionales para la publicación de datos abiertos, y lo plasmó en el “Convenio Marco de Cooperación y Asistencia Técnica Interjurisdiccional de ‘Datos Judiciales Abiertos’”, que actualmente alimenta el portal de datos abiertos judiciales nacional. En el orden subnacional se lanzó recientemente el Laboratorio de Justicia Abierta e Innovación dependiente del Consejo de la Magistratura de la Ciudad de Buenos Aires. Este surgió como un compromiso en el marco del Plan de Acción Subnacional del Gobierno de la Ciudad de Buenos Aires.

En Costa Rica se destaca la iniciativa de Justicia Abierta del Poder Judicial, que promueve un conjunto de acciones a implementar en conjunto entre los funcionarios judiciales y la sociedad civil. Para ello se desarrolló un diagnóstico y un plan de acción con un formato similar al planteado por la Alianza para el Gobierno Abierto, asumiendo compromisos, estableciendo plazos para su cumplimiento y estableciendo responsabilidades sobre cada etapa del mismo.

### Introducción

Podríamos sintetizar este capítulo con la siguiente frase: no todos los datos públicos son publicables. En este apartado, se explorarán diversos conceptos asociados a derechos que en algunos casos entran en tensión con el acceso a la información pública, y protegen la divulgación de ciertos activos de información, así como una aproximación a los debates éticos respecto a los usos que se les pueden dar.

Las fronteras de la apertura de datos están determinadas por un conjunto de normativas cuyo fin es el de preservar la publicación de ciertos activos de información, aunque estos estén gestionados y almacenados en sistemas de información públicos. Sin embargo, la aparición de nuevas tecnologías y técnicas de tratamiento de estos datos generan nuevas asimetrías entre quienes poseen las herramientas y los sujetos cuyos datos son sujeto de estas aplicaciones. La inteligencia artificial y otras prácticas de explotación de grandes volúmenes de datos, que surgen como producto de la digitalización de la gran mayoría de servicios de información, crean la necesidad de discutir los límites éticos.

Casos como el de la filtración de datos de usuarios de la plataforma Facebook, que luego fueron utilizados por la firma Cambridge Analytica para generar perfiles psicográficos de votantes en diferentes elecciones en el mundo, demuestran que el uso de datos personales puede ser explotado con fines muy distintos a aquellos que los propios usuarios admiten a través de los términos y condiciones de las plataformas. Diversas normativas en el mundo intentan avanzar en la regulación de estos temas. Uno de los mejores ejemplos es el Reglamento General de Protección de Datos (GDPR por su sigla en inglés - General Protection Data Regulation), aprobado en 2016 por el Parlamento Europeo, que reemplaza las leyes de los países sobre la protección de datos personales. Como consecuencia de la aparición del GDPR surgen diversos desafíos para las empresas y actores públicos que gestionan datos, relacionados con sus técnicas de procesamiento y las prácticas de almacenamiento y mantenimiento de las bases de datos. Además fuerza la actualización de los términos y servicios de una gran cantidad de plataformas digitales y productos intensivos en el uso de datos a nivel global.

### 2.1 Ética de datos

La ética de datos se refiere a su uso y reutilización responsable y sustentable, por parte de actores públicos o privados. Responde a un conjunto de principios que contemplan el respeto de los derechos humanos y de la privacidad de los individuos, a través de procesos transparentes respecto de la gestión y manipulación de los datos personales, ya sea de forma individual o agregada.

El Reino Unido avanzó en la definición de siete principios a implementar en el marco de la administración pública a través de su “Marco de Trabajo de Ética de Datos”:

1. Se deben tener claras las necesidades del usuario y el beneficio público. Trabajar con datos de forma innovadora presenta un gran potencial para la prestación de servicios públicos, pero se deben plantear los objetivos de cada acción de forma clara, y sus implicancias para los ciudadanos y los servidores públicos.
2. Ser consciente de la normativa relevante y los códigos de conducta. Se deben comprender de forma exhaustiva los marcos regulatorios, normativas vigentes y códigos que definen las buenas prácticas y marcan los límites de lo permitido. En caso de duda, consultar con expertos.
3. Los datos a utilizar deben ser proporcionales a las necesidades del usuario. El uso de los datos debe ser proporcional a la solución necesitada por el ciudadano y se debe utilizar la menor cantidad de datos posibles con el fin de minimizar tensiones.
4. Comprender los límites de los datos. Los datos utilizados para el diseño de políticas y servicios públicos por parte del gobierno deben ser interpretados en profundidad y de forma correcta. Es esencial considerar los sesgos y límites interpretativos de los datos al momento de evaluar su incorporación como parte de un diseño.
5. Utilizar prácticas robustas y que estén al alcance del funcionario público. Los resultados que surjan de la incorporación de nuevas herramientas tecnológicas están determinados por la calidad de los datos y las prácticas utilizadas para tratarlos. El utilizar técnicas modernas sin comprender cabalmente su uso puede resultar en resultados sesgados y contraproducentes.
6. El trabajo debe ser transparente y capaz de rendir cuentas. Se debe ser transparente sobre las herramientas, datos y algoritmos utilizados para realizar este tipo de tareas, utilizando técnicas transparentes que le permitan al ciudadano auditar estos procesos.
7. Se debe ser responsable en la inclusión de datos. Es esencial que exista un plan definido respecto de la incorporación de fuentes de datos y la exposición de los mismos de forma responsable. Esto implica que tanto los usuarios finales como los que trabajen en los productos intermedios, sepan los límites y responsabilidades asociados a estar expuestos a estos datos y su reproducción.

## 2.2 Los límites de la apertura de datos

Se pueden describir grandes conjuntos de datos que conforman las fronteras entre lo público y lo publicable, algunos de ellos por cuestiones individuales y otros por asuntos colectivos. Se distingue en este sentido que se hace referencia a público al estar estos datos gestionados por funcionarios y servidores públicos, pero no es potestad del gobierno publicar estos tipos de datos, dado que por más que su gestión, almacenamiento y procesamiento sea público, los activos de información están resguardados.

Por nombrar algunos ejemplos:

- Datos referidos a estrategias de defensa nacional
- Datos personales
- Datos de estrategias de litigios públicos
- Secreto fiscal
- Secreto comercial o corporativo
- Secreto estadístico
- Datos de inteligencia

En el siguiente apartado se harán algunas aclaraciones sobre los datos personales y la privacidad, sin embargo, el resto de los conjuntos de activos de información mencionados son igualmente importantes, si estos están delimitados como datos privados en los cuerpos normativos vigentes de cada país. Los principios de apertura y los equipos dedicados a la gestión de sus políticas, deben necesariamente considerar estas cuestiones al diseñar sus iniciativas. La apertura de datos públicos, no debe chocar con la normativa vigente y por ello la presencia de asesores legales es clave para garantizar la sostenibilidad de estas políticas, sin que sean consideradas un riesgo por otras áreas gubernamentales.

### 2.3 Derecho a la privacidad y protección de datos personales

La privacidad es un derecho fundamental consagrado en diversos tratados de derechos humanos a nivel internacional. Es crucial para la protección de la dignidad humana y constituye una de las bases fundamentales de las sociedades democráticas. Es un pilar que habilita el ejercicio de los derechos de libertad de expresión, información y asociación. En este mismo sentido la Protección de Datos Personales (PDP) es una de las formas de garantizar en términos prácticos el derecho a la privacidad en el contexto de sociedad de la información, donde el almacenamiento de datos personales está en manos de actores públicos y privados.

Existe un conjunto de datos denominados “sensibles” que deben ser protegidos en todo caso, salvo que hayan sido expresamente cedidos por su titular. Son datos sensibles aquellos que de divulgarse de manera indebida, afectarían la esfera íntima del individuo o podrían ser motivo de discurso de odio o discriminación por origen racial o étnico, estado de salud, información genética, creencias religiosas, filosóficas o morales, afiliación sindical o partidaria, opiniones políticas y preferencias sexuales. Ejemplos de estos conjuntos de datos, son el padrón de afiliados partidarios, las historias clínicas de los hospitales públicos y los registros de empleados públicos. En caso de que sea necesario divulgar esta información, existen técnicas de “anonimización” que permiten la agregación de estos datos y la eliminación de ciertos indicadores que permitan su trazabilidad. Por ejemplo, se podrían publicar análisis estadísticos sobre estos datos, garantizando que no exista forma de rastrear a los individuos que componen los segmentos divulgados.

## URUGUAY

En el caso de Uruguay según la Ley de Acceso a la Información Pública N° 18.381 establece, en aras de la transparencia, que ciertos datos de los funcionarios públicos deben ser publicados, así como brindados a quienes lo soliciten. Por otra parte, según lo dispuesto en la Ley de PDP de Uruguay, la Ley N° 18.331, y en general en todas leyes de PDP de la región, existe un elenco de datos que deben ser considerados sensibles porque abarcan el fuero íntimo y más privado de la persona: se trata de aquellos datos personales que revelen origen racial y étnico, preferencias políticas, convicciones religiosas o morales, afiliación sindical e informaciones referentes a la salud o a la vida sexual.

Los Órganos de Control de Uruguay, tanto la Unidad Reguladora y de Control de Datos Personales (URCDP), como la Unidad de Acceso a la Información Pública (UAIP) coinciden en que determinados datos de los funcionarios públicos deben ser difundidos y entregados en caso de una solicitud de acceso a la información, sin necesidad de solicitar el consentimiento informado del titular. En tanto que algunos de ellos, también pueden ser parte de la transparencia activa del organismo sin necesidad del consentimiento.

Los datos que deben ser brindados o publicados en su caso, serían aquellos que hace a la función o cargo que se posee en la administración pública (cargo, sueldos, viáticos que perciben, formación, entre otros relativos al cargo o función pública).

Si todos estos datos tuvieran el carácter de “datos sensibles”, para brindar acceso a los mismos la administración debería recabar siempre el consentimiento informado del titular, ya que es una característica fundamental del carácter del dato sensible. Por ende, esclarecer este punto es esencial para trabajar el concepto de Gobierno Abierto y el derecho de acceso a la información pública, porque involucra el principio de transparencia y rendición de cuentas que debe guiar el accionar de la administración pública.

\*Agradecemos las especificaciones sobre Uruguay a la Dra. Graciela Romero - Unidad de Acceso a la Información Pública

## ARGENTINA

En la República Argentina existen diversos casos en los que se han debatido los límites sobre qué tipo de datos son alcanzados por las normativas específicas que los protegen y en cuáles de estos casos colisionan con el derecho de AIP y el de PDP. En el caso Garrido contra AFIP la Corte Suprema de Justicia de la Nación estableció jurisprudencia clara sobre el marco en el cuál todos aquellos datos que nos son “sensibles” deben ser provistos en caso de ser requeridos a través de los mecanismos previstos en los principios de AIP y sus respectivas normativas. Un ejemplo innovador es el planteamiento institucional que se le ha dado a la Agencia de Acceso a la Información Pública del Poder Ejecutivo, el organismo encargado de velar por el cumplimiento efectivo de la Ley 27275/16 en el contexto del Poder Ejecutivo de la Nación Argentina, dado que se le atribuyeron las funciones de la ex-Dirección Nacional de Protección de Datos Personales que hasta 2017 residía en el marco del Ministerio de Justicia de la Nación. Esta agencia tiene una autonomía relativa respecto del ejercicio de su rol, y le toca velar por el cumplimiento de dos derechos que en algunos casos se encuentran en tensión.

En caso de que sea necesario divulgar esta información, existen técnicas de “anonimización” que permiten la agregación de estos datos y la eliminación de ciertos indicadores que permitan su trazabilidad. Por ejemplo, se podrían publicar análisis estadísticos sobre estos datos, garantizando que no exista forma de rastrear a los individuos que componen los segmentos divulgados.

### 2.4 Las fronteras de la discusión

Un eje que cada vez se vuelve más central en la discusión sobre el uso ético de los datos, es la aparición de tecnologías que a través de algoritmos y técnicas de machine learning toman decisiones sobre personas. Machine learning, o aprendizaje automático, es la denominación utilizada para identificar diversos conjuntos de técnicas a través de las cuáles se le “enseña” a computadoras a optimizar procesos de toma de decisiones, en muchos casos basados en decisiones previas tomadas por humanos, en algún problema en particular. Este tipo de prácticas han sido utilizada en las últimas décadas por el sector privado en diversas áreas, como la industria financiera y bancaria. Las “redes sociales” como Facebook o Twitter implementan estas técnicas diariamente para optimizar el alcance

de sus contenidos y Google las utiliza para mejorar los resultados de las búsquedas.

El foco que une esta conversación con las políticas de gestión de datos, es visible en el momento en que los gobiernos comienzan a incorporar estas técnicas avanzadas de implementación de algoritmos para la toma de decisiones en procesos de política pública. Existen diversos casos en el mundo y la región en los que se han tomado decisiones a través de herramientas tecnológicas, que han reproducido sesgos pre-existentes, así como casos en donde la tecnología utilizada no era comprendida cabalmente por los servidores públicos y luego no podían explicar los resultados obtenidos.

ILDA ha publicado recientemente un análisis sobre el uso de estas tecnologías en América Latina denominado “Automatizar con cautela: Datos e Inteligencia Artificial en América Latina”. Este documento se centra en dos casos principalmente, uno de Uruguay y otro de Argentina, y fue el eje de un taller realizado en Montevideo en el que participaron diversos actores de gobiernos, sociedad civil y academia de la región para debatir sobre los riesgos y alcances de la temática.

Por ello, es necesario ampliar la conversación de la transparencia y la rendición de cuentas de datos e información pública, hacia la transparencia y rendición de cuentas de los algoritmos utilizados en procesos públicos de toma de decisión que afecten de forma directa o indirecta a la ciudadanía. Es una discusión compleja y todavía no está del todo claro cómo se puede saldar en el corto plazo, pero es central que se contemplen estos debates en el futuro.

# Licenciamiento de la información.

## Introducción a las licencias para uso de datos abiertos

### Introducción

Para que un dato sea considerado efectivamente abierto debe cumplir con dos condiciones fundamentales: estar publicado en formatos abiertos y poseer una licencia que de forma expresa habilite su reutilización. Esta unidad se enfocará en modelos de licenciamiento que puedan ajustarse a las políticas de datos abiertos.

Las políticas de apertura de datos deberían tener como eje central el fomentar su reutilización por parte de diversos actores de la comunidad. Como primer paso, se deben establecer condiciones claras sobre cuáles son las posibilidades y límites de uso de los datos, para los usuarios que quieran agregarle valor. Esto se determina explícitamente en los términos y condiciones de uso de los datos, lo cual en la gran mayoría de los casos, se expresa en las licencias de uso a las cuáles están atadas los activos de información y las plataformas de publicación.

### 3.1 Derecho de propiedad intelectual

Esta conversación se ancla en los límites del derecho de propiedad intelectual, reconociendo en su seno el derecho de las personas físicas y jurídicas a recibir un rédito económico por sus creaciones. Ejemplos de registros públicos que asignan esta propiedad a sus autores son las patentes, derechos de autor, derechos de marca, denominaciones de origen, entre otros.

Es importante distinguir que los bienes, productos y servicios prestados por el estado configuran una figura particular que es la de “bienes públicos”. Este tipo de bienes son propiedad del estado y sus derechos intelectuales le corresponden al estado en su conjunto. Las bases de datos y activos de información pueden considerarse de esta forma, y solamente es necesario distinguir los usos permitidos a través de un licenciamiento explícito en el soporte en el cual estén publicados.

El derecho de autor (copyright) es la forma en la cual se expresa el derecho de propiedad intelectual, demostrando que los contenidos expuestos junto al símbolo © tienen los derechos de reproducción reservados, es decir, no se pueden copiar, almacenar, distribuir ni proyectar sin la autorización escrita de quien posea los derechos de autor, lo que puede incluir el pago por el uso de los contenidos. En un gran conjunto de sitios web de gobierno es común encontrarse con estas leyendas que aclaran que los derechos están reservados, lo cual contradice el sentido público de la información y de los datos allí publicados. En muchos casos esto responde a un desconocimiento sobre las implicaciones reales sobre el copyright y las licencias abiertas y el dominio público, y en otros a decisiones que toman los proveedores de las páginas web y no los funcionarios responsables. Es importante entonces concientizar a los funcionarios de los gobiernos sobre los tipos de licencias que se exhiben en sus sitios web, y exhortar a que se promueva la reutilización de los datos públicos de forma explícita a través de licencias que permitan su uso.

### 3.2 ¿Licencias abiertas o cerradas?

En paralelo con el desarrollo de internet surgieron diversas corrientes y comunidades con énfasis en la idea de que el conocimiento de la humanidad es propiedad de la humanidad en su conjunto. Esto se contrapone con los planteamientos descritos anteriormente sobre la propiedad intelectual dado que al considerarse este tipo de aportes bienes comunes, se contradice con la visión de lucro que motiva a los derechos de autor.

La comunidad de código abierto, que promueve el uso de software desarrollado de forma colectiva ha sido una de las grandes promotoras de estas discusiones. Bajo esta lógica surgió el concepto de crear un nuevo tipo de licencia que permita estas prácticas sin quedar por fuera de los marcos legales mayormente aceptados en nuestros países. Así surgieron las licencias Creative Commons que se describen a continuación.

#### Creative Commons

Las licencias Creative Commons están compuestas por cuatro grupos de condiciones particulares que el creador le puede asignar a su producto:

- **Atribución:** en cualquier explotación de la obra autorizada por la licencia será necesario reconocer la autoría (obligatoria en todos los casos).
- **No comercial:** la explotación de la obra queda limitada a usos no comerciales.
- **Sin obras derivadas:** La autorización para explotar la obra no incluye la posibilidad de crear una obra derivada.
- **Compartir igual:** la explotación autorizada incluye la creación de obras derivadas siempre que mantengan la misma licencia al ser divulgadas.

En la siguiente lista se describen las licencias que surgen de la combinación de las condiciones previamente descritas:



**Atribución (by):** Se permite cualquier explotación de la obra, incluyendo la explotación con fines comerciales y la creación de obras derivadas, la distribución de las cuales también está permitida sin ninguna restricción. Esta licencia [es una licencia libre según la Freedom Defined](#).



**Reconocimiento - Compartir Igual (by-sa):** Se permite el uso comercial de la obra y de las posibles obras derivadas, la distribución de las cuales se debe hacer con una licencia igual a la que regula la obra original. Esta licencia [es una licencia libre según la Freedom Defined](#).



**Atribución - No Comercial (by-nc):** Se permite la generación de obras derivadas siempre que no se haga con fines comerciales. Tampoco se puede utilizar la obra original con fines comerciales. Esta licencia no es una licencia libre.



**Atribución - No Comercial - Compartir Igual (by-nc-sa):** No se permite un uso comercial de la obra original ni de las posibles obras derivadas, la distribución de las cuales se debe hacer con una licencia igual a la que regula la obra original. Esta licencia no es una licencia libre.



**Atribución - Sin Obra Derivada (by-nd):** Se permite el uso comercial de la obra pero no la generación de obras derivadas. Esta licencia no es una licencia libre.



**Atribución - No Comercial - Sin Obra Derivada (by-nc-nd):** No se permite un uso comercial de la obra original ni la generación de obras derivadas. Esta licencia no es una licencia libre, y es la más cercana al derecho de autor tradicional.

Fuente: Creative Commons Argentina

## Otros ejemplos de licencias abiertas

Se describen aquí otros casos y ejemplos de licencias abiertas, considerando tanto esfuerzos internacionales de la sociedad civil y de gobiernos que han desarrollado sus propios marcos normativos con foco puesto en la información pública:

- **Dominio público** es el ejemplo más claro sobre una licencia abierta, que en realidad toma efecto al caducar una licencia de derecho de autor. En este sentido hay diferentes normativas alrededor del mundo, pero en términos generales luego de cumplidos 70 años (para obras artísticas, literarias, musicales y fílmicas) o 50 años (para radiodifusiones y grabaciones musicales), los dueños de los derechos dejar de cobrar regalías por los mismos y las obras pasan a ser parte del dominio público.
- El Reino Unido creó la “Open Government Licence for public sector information” a través de sus Archivos Nacionales. Esta licencia permite la copia, publicación y transmisión de información, la readaptación de los datos y la posibilidad de explotar comercialmente y no comercialmente los datos y la información, a través de la combinación con otras fuentes o incluyendo los datos e información pública en productos o servicios desarrollados por terceros.
- La Agencia de Gobierno Electrónico y Sociedad de la Información y del Conocimiento (AGESIC) de Uruguay ha establecido una Licencia de Datos Abiertos propia, a través del art 82 de la Ley 19355, como licencia obligatoria para todos los organismos del Estado que publiquen datos abiertos.
- Open Data Commons Open Database License (ODbl) es una licencia desarrollada por la Fundación Internacional Open Knowledge, con foco puesto específicamente en la publicación de conjuntos de datos abiertos a través de portales ad-hoc.

### Introducción

Los Datos Abiertos Gubernamentales (DAG) son parte de una familia mayor, la de los Datos Abiertos. Aunque en general se utilizan ambos términos de forma equivalente, los Datos Abiertos pueden tener diversos orígenes. En esta unidad incorporaremos diferentes conceptos alrededor de los Datos Abiertos. Según la Carta Internacional de Datos Abiertos: “Los Datos abiertos son datos digitales que son puestos a disposición con las características técnicas y jurídicas necesarias para que puedan ser usados, reutilizados y redistribuidos libremente por cualquier persona, en cualquier momento y en cualquier lugar”. La Carta es un esfuerzo que surge de la conversación entre gobiernos y sociedad civil y promueve la adopción de los seis principios que se describen a continuación.

### 4.1 Carta por los Datos Abiertos

La Carta Internacional de Datos Abiertos (Open Data Charter en inglés) es el producto de una colaboración entre gobiernos y organizaciones que trabajan en la apertura de datos basados en una serie de principios. Estos principios son de naturaleza aspiracional y se espera que los gobiernos los adopten como referencia de mejores prácticas. Sin embargo, los principios ni los documentos del Open Data Charter son vinculantes para los gobiernos bajo normativa internacional ni nacional.

El objetivo del Charter es promover la cultura abierta y transformarlas en prácticas organizacionales de manera que las mismas sean resilientes a los cambios de gobiernos. Para esto, existe un proceso de “adopción” del Open Data Charter en el cual un gobierno debe realizar un compromiso político de alto nivel con los principios anteriormente citados, definiendo metas y entregables.

A marzo del año 2019, 69 gobiernos a distintos niveles han adoptado el Open Data Charter, además, 49 organizaciones y actores no gubernamentales han públicamente avalado el Charter. Notoriamente, la Alianza para Gobierno Abierto (Open Government Partnership) no está listada como una de las organizaciones que avale el Open Data Charter.

Los principios establecen que la publicación de los datos deben tener las siguientes características[1]:

1. **Abierto por defecto (open by default):** este principio significa que los datos del gobierno deben estar públicamente disponibles para que cualquiera pueda encontrarlos, acceder a ellos y utilizarlos bajo una licencia abierta y no restrictiva, a menos que exista una razón legal que lo impida, en cuyo caso dicha razón será comunicada públicamente.
2. **Oportuna y completa (timely and comprehensive).** Los datos se deben publicar tan pronto sea posible, y con actualizaciones periódicas de acuerdo a la naturaleza del dato. Considerando que existen recursos limitados para realizar tareas de publicación de datos por el gobierno, se reconoce la importancia de realizar consultas ciudadanas con el fin de priorizar los conjuntos de datos a ser publicados, para asegurar su uso.

Los datos publicados deben ser completos y de buena calidad.

3. **Accesible y utilizable (accessible and usable).** Una vez que los datos son publicados, deben ser fácilmente encontrables y accesibles, sin barreras administrativas ni burocráticas para su acceso, sin costos monetarios, e idealmente, sin necesidad de registro como usuario en el sistema. Además, para fomentar el uso de los datos, se reconoce la importancia del uso de los formatos de datos abiertos (ver Unidad 2).
4. **Comparable e interoperable:** Se reconoce el valor del uso de estándares de datos, lo que permitiría que los datos sean comparables en varios ejes (temporal, geográfico, por sectores o proyectos). El uso de estándares permite también la integración de datos para su análisis. (ver Unidad 5).
5. **Para mejorar la gobernanza y la participación ciudadana.** Se reconoce que la publicación de los datos abiertos fortalece la gobernanza y la confianza en las instituciones públicas, mejorando la capacidad de rendición cuentas y el control por la ciudadanía, lo que se traduce en mejores servicios y programas para todas las personas. Se promueven los procesos de consultas e interacción con los ciudadanos, especialmente para los procesos de priorización de los datos a ser publicados.
6. **Para el desarrollo inclusivo y la innovación.** El uso de los datos abiertos puede promover la innovación en la sociedad, sector privado, academia y gobierno. Esta innovación a su vez se puede traducir en beneficios económicos y sociales. El uso de procesos de innovación abierta y participativa puede también ayudar a encontrar soluciones colaborativas y problemas de alto interés e impacto social y económico.

En relación con el cuarto principio planteado por la carta, el Manual de Datos Abiertos hace hincapié en la relación entre los conceptos de apertura e interoperabilidad. Es decir, la habilidad de diversos sistemas y organizaciones para trabajar juntos (interoperar). En este caso, se refiere a la habilidad para interoperar o integrar diferentes conjuntos de datos.

El concepto de “Datos Abiertos” ha pasado de ser un concepto inexistente, hasta hace unos pocos años, a convertirse en una agenda de trabajo importante para funcionarios de gobierno, emprendedores y activistas. Esta agenda sería impensable sin el desarrollo de las TIC, ni el trabajo previo de los grupos de activistas de AIP (a nivel internacional y nacional), principalmente a finales de los años 90 y principios de la década de 2000.

#### 4.2 El proceso de apertura de datos y sus etapas de valor

Aquí nos centraremos en comprender los procesos de valor agregado que sustentan el desarrollo de las políticas. La cadena de valor parte desde el dato como una externalidad económica de la digitalización de los procesos de la gestión pública y privada, e implica necesariamente un conjunto variado de actores, interacciones, tecnologías e intervenciones para que se pueda desarrollar.

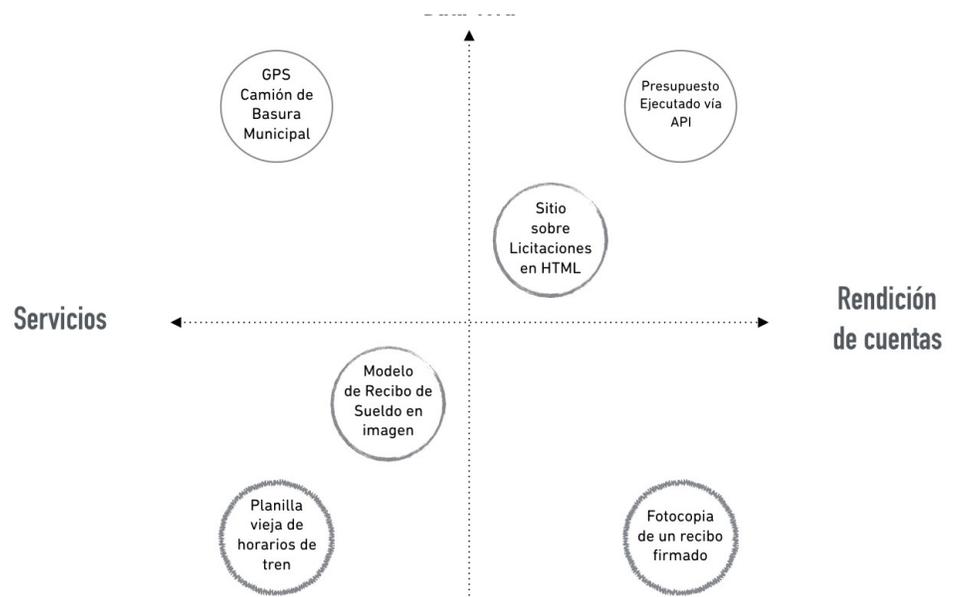


Fuente: Belbis, 2013.

### 4.3 Almacenamiento y generación de los datos

Los datos están ahí. En servidores físicos dentro de las organizaciones o en la nube, pero el almacenamiento de información estructurada o no estructurada se da desde el momento en que se comenzaron a normalizar como estándares de diferentes industrias, la digitalización de sus procesos de gestión, stock, finanzas, entre otros. A esto se suman empresas u organizaciones centradas específicamente en información digital, como las redes o servicios sociales, sitios web, consultoras de datos, organismos internacionales.

En esta etapa del proceso se implementan, en términos tecnológicos, diferentes tipos de bases de datos, servidores físicos y numerosos sistemas de gestión documental tanto analógicos como digitales. Yu y Robinson (2012), construyen una clasificación que ubica a la información gubernamental en un eje según sea inerte o adaptable. La información inerte es aquella que por su formato de origen no digital implica altos costos para ser convertida en datos abiertos. En el otro extremo se encuentra aquella que ya está digitalizada y requiere simplemente un proceso de refinamiento y puesta en disponibilidad, que depende de voluntad política.



Fuente: Elaboración propia basada en Yu & Robinson (2012).

### 4.4 Refinamiento y limpieza

En el refinamiento de los datos, la segunda etapa de esta cadena de valor, los actores involucrados son menos y de mayor especificidad técnica y profesional. La figura del Data Miner es central en este punto ya que puede aplicar sus conocimientos y técnicas para acelerar los procesos de limpieza de las bases de datos y construir las estructuras informáticas necesarias para su almacenamiento y difusión.

Estos procesos se plantean como funciones de las oficinas de sistemas internas de cada organización, o se pueden tercerizar e integrar a servidores externos o servicios en la nube. En esta etapa se emplean tecnologías de gestión de bases de datos, análisis estadístico, análisis de procesos y diseño de información.

## 4.5 Portales y publicación

Una vez que tenemos los datos y avanzamos en su refinamiento para que sean liberados al público nos encontramos con diferentes opciones. Existen hoy desarrollos tecnológicos específicos para la publicación de datos abiertos –CKAN, Socrata, Junar- que permiten la creación de catálogos de datos en línea, disponibles a través de un API y la personalización de la interfaz de usuario. Además de estos portales de datos, en algunos casos particulares existen desarrollos hechos ad-hoc para ciertas iniciativas.

## 4.6 Reutilización

Hasta aquí logramos que los datos que ya estaban generados, hayan sido limpiados, refinados y publicados en un portal que permita su reutilización. ¿Y ahora?

Este es el punto más crítico de la cadena de valor, dado que si no se genera algún resultado concreto a partir de este punto, todos los esfuerzos se truncan. De nada sirve tener una plataforma con decenas, cientos o miles de conjuntos de datos si nadie los reutiliza para generar productos, servicios, emprendimientos, etc. La potencialidad debe convertirse en valor concreto.

## 4.7 El desafío de la reutilización

Beth Noveck (2012) plantea que la apertura de datos se convierte en la base de la innovación pública fundamentalmente porque estos datos sirven como canal de vinculación con agentes privados y públicos, para integrarlos en la resolución de problemas complejos dentro de las esferas gubernamentales. Esta tendencia de liberar datos de gobiernos locales, regionales y nacionales le permite a la ciudadanía en su conjunto, ya sea desde los individuos, desde organizaciones privadas, o desde el mismo sector público, incurrir en su reutilización para la generación de nuevos productos e iniciativas (Kulk & Loenen, 2012).

La idea de Noveck sirve para ir un poco más allá de una primera concepción intuitiva que ata directamente a los datos abiertos como una solución a la transparencia, y nos abre todo un nuevo campo de acción bajo el concepto de la innovación pública. Son varias las expresiones actuales de la innovación en términos gubernamentales: laboratorios dentro del gobierno, hackatones, colaboración con el tercer sector en proyectos. Pensar entonces los datos abiertos gubernamentales como la materia prima sobre la cual elaborar y desarrollar soluciones, políticas públicas y productos estatales es un desafío interesante para madurar el proceso de Gobierno Abierto.

Con estas ideas y parámetros, se plantean un conjunto de desafíos para avanzar con los procesos de apertura de datos públicos en formatos que sean reutilizables. La reutilización de estos datos genera un campo fértil sobre el cual las organizaciones de la sociedad civil, los miembros del sector público y la ciudadanía se paran de cara a la generación de innovación cívica.

## 4.8 El rol de los intermediarios

La reutilización no es producto sólo de los esfuerzos gubernamentales, sino que surge de un conjunto de actores de la sociedad civil y del sector privado que presentan habilidades que los ponen en el centro de esta conversación.

**Sociedad civil:** el tercer sector es en muchos casos el primer actor en vincularse con la agenda de datos abiertos, probablemente más inclinados sobre el eje de la transparencia y la rendición de cuentas. La reutilización por parte de estos actores tiene implicaciones directas en procesos de escrutinio público, incidencia, así como el desarrollo de aplicaciones o plataformas de participación ciudadana en nichos particulares.

**Periodismo:** el sector periodístico tiene diferentes enfoques en la reutilización de datos abiertos. Existen medios que conforman sus equipos específicos de periodismo de datos que con equipos multidisciplinarios encaran proyectos de visualizaciones interactivas e investigaciones centradas en datos. Aún en los medios que no dedican recursos exclusivos para este tipo de emprendimientos en sus redacciones, el uso de los datos abiertos facilita investigaciones más profundas.

**Academia:** el sector académico se convierte en un reutilizador al usar los datos abiertos como objeto de estudio. Pero el verdadero aporte de los datos abiertos a los procesos de enseñanza, es como recursos educativos que puedan servir en contextos de problemáticas públicas concretas para aportar soluciones innovadoras.

**Emprendedores tecnológicos:** dentro del sector privado, aquellos que se suelen acercar en principio a los espacios de innovación cívica y reutilización de datos son los emprendedores tecnológicos, ya que los datos son materia prima para la innovación. En este sentido estos actores tienen habilidades propias y conocimiento técnico que permiten la creación de aplicaciones móviles, servicios digitales y plataformas tecnológicas intensivas en datos.

**Servidores públicos:** el quinto conjunto de actores está compuesto por los propios empleados y funcionarios que componen las diferentes instituciones públicas. El uso y la generación de datos públicos es un proceso complejo que suele estar disperso en gran parte de las áreas operativas que conforman un gobierno. La difusión de las prácticas de apertura de datos es central para evitar duplicidades en la construcción de indicadores, análisis de políticas públicas y registros que además, en su gestión no centralizada, pueden generar confusiones, errores y malas interpretaciones de datos públicos.

Esta enumeración de actores reutilizadores no es exhaustiva, y existen diversos ejemplos de otros actores de la sociedad civil y el sector privado que se han involucrado en las agendas de datos abiertos y utilizado estos datos para la generación de nuevos productos, servicios y políticas públicas.

### Introducción

AIP y DAG son dos maneras de abordar la divulgación de información y datos producidos por los gobiernos. En esta unidad nos enfocaremos en el derecho de Acceso a la Información Pública introduciendo algunas ideas sobre Datos Abiertos.

Desde fines del siglo XX, el derecho de acceso a la información pública (AIP) es reconocido como un derecho humano por diversos organismos internacionales. Durante la primera década del siglo XXI, el AIP comenzó a experimentar transformaciones, como consecuencia de la evolución de las Tecnologías de Información y Comunicación (TIC). Es decir, comenzó a expandirse y abarcar no solo a los documentos oficiales sino también a la información contenida en diferentes formatos digitales (correos electrónicos, por ejemplo). Asimismo, se comenzó a pasar de la divulgación mayormente reactiva (a través de solicitudes según los procedimientos establecidos por la legislación de AIP) a una combinación entre el acceso a la información vía solicitudes y la divulgación proactiva de información (principalmente a través de los sitios web gubernamentales). En paralelo, se ha ampliado el alcance del acceso para abarcar también datos en formatos digitales abiertos, conocidos como Datos Abiertos Gubernamentales (DAG).

### 5.1 Acceso a la información pública

El derecho de acceso a la información (AIP) se entiende como la capacidad de solicitar y obtener información producida y resguardada por los gobiernos, con sujeción a las excepciones definidas en la regulación en la materia. Desde la promulgación de la primera regulación relacionada con la posibilidad de acceder a información en manos del Estado, en Suecia en 1766, hasta finales del siglo XX, el AIP no fue un tema relevante dentro de las agendas de política pública en la gran mayoría de los países. Sin embargo, entre los primeros años de los noventa y finales de los 2000, el tema atrajo mucha atención y fue denominado por Darch and Underwood (2010) como la "Edad de Oro". Durante esos años se observó una gran expansión del número de leyes que regulan el AIP; más de 70.

Durante este período, dos procesos interrelacionados fueron extremadamente significativos en cuanto a la difusión de la legislación de AIP a nivel global: la internacionalización del concepto y los estándares en la materia, y el trabajo de los activistas de sociedad civil (Fumega 2015). Unos años después, durante los primeros años del siglo XXI, el impacto de las TIC se hizo evidente cuando una nueva agenda cobró popularidad en el área de información gubernamental: Los Datos Abiertos Gubernamentales (DAG).

## 5.2 Debate entre las dos agendas

Los datos son la unidad mínima que procesada puede convertirse en información. De todos modos, esta distinción semántica no es la única divergencia entre las dos agendas: AIP y DAG.

La legislación de AIP permite a los individuos (principalmente a través de una solicitud) acceder a la información en posesión de los poderes del Estado mientras que las políticas de DAG proactivamente proporcionan a las personas datos (el componente primario necesario para producir información) en formatos que permitan su reutilización. Por lo tanto, las políticas de DAG enfatizan en el contenido de lo publicado y en sus formatos.

Los desarrollos basados en las TIC han transformado todos los ámbitos relacionados con la información, desde la gestión, su uso/reutilización y su almacenamiento. La influencia de las TIC puede canalizarse y entenderse de dos maneras. Por un lado, la influencia más clara se relaciona con la mejora en los métodos de comunicación y de gestión de la información. Por otra parte, con la influencia que ejercen los profesionales que trabajan en el área. En muchos casos, su visión se refleja en su trabajo, en la manera de relacionarse y en la manera en que estructuran sus organizaciones.

En otras palabras, los desarrollos de las TIC inciden en los campos relacionados con la información porque proporcionan nuevas herramientas para su gestión y porque la ética/cultura hacker, entre otros aspectos filosóficos, proporcionan una "visión del mundo" particular, diferente al enfoque más legalista de derechos. De esta manera, la división fundamental entre estas dos áreas se deriva de las herramientas y conocimientos específicos de cada una de ellas, y de las influencias ligadas a las profesiones más preponderantes dentro de cada una de las agendas (Fumega, 2016).

En este contexto, las organizaciones que trabajan en los campos de legislación de AIP y de políticas de DAG si bien comparten los recursos primarios- información y datos producidos y resguardados por las agencias gubernamentales- difieren en las influencias filosóficas y profesionales con las agrupaciones de AIP. Estas últimas se enfocan mayoritariamente en el acceso mientras que las de DAG se orientan más hacia el uso y el valor que puedan agregar a los datos (Fumega, 2016).

Estas diferencias explican, en gran parte, la división entre las dos agendas. Con los años, la brecha entre ambas agendas se ha reducido. De todos modos, todavía falta mucho para que se piense a ambas agendas como aspectos de un mismo ecosistema de flujos de información y datos.

En relación con la proactividad expresa en las dos políticas públicas, entendiendo que ambas tienen a la transparencia como un eje central, es posible plantear al AIP en términos de transparencia pasiva, garantizándole al ciudadano la capacidad de solicitar información pública y obligando al funcionario público a brindarla. Por otro lado, los DAG son el lado activo, en el cual se obliga a los gobiernos a publicar ciertos activos de información sin que sea el ciudadano el que ejecute la acción de solicitarlos.

### 5.3 Políticas de Datos Abiertos a nivel comparado

En la región se pueden destacar cinco casos de políticas de apertura de datos: México, Uruguay, Brasil, Argentina y Chile. De acuerdo a la última edición del Barómetro de Datos Abiertos de 2017, estos cinco países se encuentran en el top 20, en los puestos 6, 11, 14, 17 y 20 respectivamente.

País	Posición	Puntaje	Preparación	Implementación	Impacto
Argentina	17	47	66	56	20
Brasil	14	50	63	56	30
Chile	20	40	54	42	30
México	6	69	79	67	62
Uruguay	11	56	71	70	28

Fuente: Open Data Barometer

En Brasil la Infraestructura Nacional de Datos (INDA), que fue parte de los primeros compromisos del país en el AGA/OGP, es un conjunto de tecnologías, procedimientos y mecanismos que aseguran las condiciones necesarias para la diseminación de información y datos públicos en formatos abiertos. Esta infraestructura se encuentra regulada por la “instrucción normativa N° 4” del 12 de abril de 2012. Más recientemente todos estos desarrollos en materia de política de datos abiertos fueron regulados por el Decreto 87775 de 2016. Una de las principales características de esta normativa es la asignación de responsabilidades a áreas específicas de la administración pública.

En México, a comienzos de 2015, se sancionó el Decreto que regula su política de Datos Abiertos. Si bien el decreto solo establece las principales definiciones de la política de datos abiertos, es un instrumento que permite sentar los principios básicos sobre los cuales se construye esa política.

En el caso de Brasil, los objetivos descritos se encuentran orientados a la promoción de la publicación y capacitación en la materia mientras México estipula que el principal objetivo de la regulación es el ordenamiento de la publicación para así facilitar su uso.

De esa manera, el plan Nacional de Apertura de Datos de Argentina se encuentra regulado por el Decreto 117/2016 que reglamenta plazos(180 días)para la publicación de las primeras bases de datos en el Portal de Datos, y para detallar los activos de datos de los distintos organismos de la Nación y su cronograma de publicación. En el mismo sentido, el artículo 32 de la Ley 27275/16 de Acceso a la Información Pública regula el principio de Transparencia Activa, definiendo un conjunto de activos de información que deben ser publicados por todos los sujetos obligados de la Ley.

Estos tres ejemplos aquí mencionados incluyen provisiones sobre la importancia de los datos abiertos para el desarrollo de diferentes áreas de gobierno y de la sociedad. Asimismo, en todos los casos se establecen responsabilidades para la coordinación de la política (dentro del Poder Ejecutivo Nacional) a una sola oficina y para la preparación y publicación de cada uno de los conjuntos de datos a las agencias que producen y resguardan esos datos. Esta responsabilidad descentralizada no sólo evita

que toda la carga de curación y publicación recaiga en manos de una sola agencia, sino que también permite que cada una de las agencias responsables generen capacidades internas para poder cumplir con la normativa. Adicionalmente, las oficinas que coordinan la política de datos abiertos se atribuyen la responsabilidad, en la mayoría de los casos, de proveer asistencia técnica en la materia. En las normativas referidas se asignan responsabilidades por el desarrollo del plan de apertura y de publicación de los datos y se establece un cronograma.

Por último, en los casos analizados las restricciones a la publicación de datos se encuentran sujetas a lo estipulado por las leyes y/o normativas que regulan el derecho de acceso a la información pública, así como otras normativas que protegen la privacidad y los datos personales, entre otros. Por este motivo, es importante que las regulaciones y políticas de datos abiertos se encuentren alineadas con la legislación vigente en términos de acceso a la información y de protección de datos personales. Por otra parte, las normativas en términos de acceso a la información deben enmendarse en el caso que, por la fecha de sanción, no contengan previsiones sobre los formatos abiertos en los cuales se debe entregar y publicar los datos.

## **La Unidad de Acceso a la Información Pública (UAIP)**

Es un órgano especializado cuyo cometido es velar por el respeto al Derecho de Acceso a la Información Pública:

- Brinda asesoramiento Jurídico al Poder Ejecutivo en materia de acceso a la información pública.
- Controla el cumplimiento por parte de los organismos públicos de las estipulaciones previstas en la Ley de Acceso a la Información Pública.
- Promueve el cumplimiento de los derechos y obligaciones, mediante la capacitación de funcionarios.
- Genera políticas que facilitan el acceso a la información y la transparencia en el Estado.
- Resuelve peticiones presentadas por las personas frente a la denegación o falta de respuesta por parte de los organismos.

La Unidad de Acceso a la Información Pública (UAIP) fue creada por el artículo 19 de la Ley N° 18.381 de Acceso a la Información Pública como un órgano desconcentrado de la AGESIC.

Puedes aprender más en el curso de la UAIP sobre la Ley de Acceso a la Información Pública.

## **La Unidad Reguladora y de Control de Datos Personales (URCDP)**

Fue creada por la Ley N° 18.331 de Protección de Datos Personales y Acción de Habeas Data (LPDP), con autonomía técnica, cuya competencia es custodiar el cumplimiento de la legislación de protección de datos personales y asegurar el respeto de sus principios.

Puedes aprender más en el curso de la URCDP sobre Protección de Datos.

## **La Unidad Reguladora y de Control de Datos Personales (URCDP)**

En enero de 2019 entró en vigencia la nueva Ley de Rendición de Cuentas, que incorpora importantes modificaciones a la legislación nacional sobre Protección de Datos Personales con el fin ofrecer mayores garantías a los uruguayos.

Dichos cambios tienen el objetivo de alinear la legislación nacional con los nuevos desarrollos en la materia, ofreciendo así mayores garantías a las personas para la protección de sus datos personales.

Los cambios a la legislación sobre Protección de Datos Personales (introducidos por los artículos 37 a 40 de la Ley de Rendición de Cuentas) son los siguientes:

- Ampliación del ámbito de aplicación de la Ley de Protección de Datos Personales
- Nuevas obligaciones para responsables y encargados de bases de datos
- Modificaciones al “principio de responsabilidad”
- Creación de la figura del “delegado de protección de datos”

## **La Unidad Reguladora y de Control de Datos Personales (URCDP)**

La Ley N° 19.355, del 19 de diciembre de 2015, estableció en su artículo 82 la obligación para las entidades públicas de publicar en formato abierto la información preceptuada por los artículos 5° y 8° de la Ley N° 18.381, del 17 de octubre de 2008, y los artículos 38 y 40 de su Decreto Reglamentario N° 232/010, del 2 de agosto de 2010. Este artículo establece lo siguiente: Las entidades públicas deberán como mínimo publicar en formato abierto, la información detallada en el artículo 5 de la Ley N° 18.381 y los artículos 38 y 40 del Decreto Reglamentario N° 232/010.

Los datos y sus metadatos asociados deberán cumplir con las normas técnicas que determine la Agencia para el Desarrollo del Gobierno de Gestión Electrónica y la Sociedad de la Información y del Conocimiento. La publicación de estos datos deberá realizarse en el marco de lo dispuesto por la Ley N° 19.179, de 27 de diciembre de 2013 ("software" libre y formatos abiertos en el Estado).

# ARGENTINA

## **El Derecho al Acceso a la Información Pública en Argentina**

La regulación del Derecho al Acceso a la Información Pública (AIP) está en un proceso de evolución y cambios en la República Argentina. En el año 2003 el Decreto 1172 estableció un régimen regulatorio sobre el derecho que sólo aplicaba al Poder Ejecutivo. En este sentido la aprobación de la Ley 27275 de 2016 fue un avance en la garantía del mismo en toda la Administración Pública Nacional y otras organizaciones de interés público, incluyendo a gremios, partidos políticos, empresas públicas y otros sujetos obligados.

Entendiendo al federalismo que prima como ordenamiento político en Argentina, tanto los municipios como las provincias son responsables de dictar sus propias normativas en pos de garantizar el derecho en sus contextos de aplicación local. En este sentido existen diversas experiencias

de municipios y provincias con decretos, ordenanzas y leyes propias que establecen los mecanismos y alcances de las mismas.

En el caso de la Provincia de Mendoza, la recientemente aprobada Ley 9070/18 establece el derecho y sus mecanismos de acceso por parte de los ciudadanos, comprendiendo entre sus sujetos obligados a:

- Poder ejecutivo
- Administración centralizada, organismos descentralizados, organismos desconcentrados, entes autárquicos, administraciones mixtas y/o cualquier otra entidad u organismo que administre recursos públicos, bienes del Estado o ejecute actos de la administración pública en general.
- Empresas y sociedades con participación del Estado, entes residuales y todas las organizaciones empresariales donde el Estado Provincial tenga participación.
- Poder Legislativo
- Poder Judicial, Ministerio Público de la Defensa y Ministerio Público Fiscal.
- Fiscalía de Estado.
- Tribunal de Cuentas.
- Departamento General de Irrigación.
- Prestatario o concesionarios de servicios públicos o que exploten un dominio público.
- Toda entidad privada o pública que por tiempo determinado, encomendada por el Estado, realiza la organización y el funcionamiento de un servicio público o explotación de un dominio público.

Existen en el orden municipal casos de relevancia como el de la Ciudad de Mendoza y el Departamento de Godoy Cruz, que poseen sus normativas locales aprobadas por los correspondientes Concejos Deliberantes.

## MÓDULO 1

# Lecturas

Fumega, S., & Scrollini, F. (2012). El acceso a la información y los gobiernos abiertos en América Latina . En Hofmann, Ramírez-Alujas, & Bojórquez, La promesa del gobierno abierto. (págs. 225-257). Ciudad de México: ITAIP. Accesible en el sitio: <http://inicio.ifai.org.mx/Publicaciones/La%20promesa%20del%20Gobierno%20Abierto.pdf#page=225>.

Hofmann, Ramírez-Alujas, & Bojórquez. (2012). Introducción. En Hofmann, Ramírez-Alujas, & Bojórquez, La promesa del gobierno abierto (págs. 19-25). Ciudad de México: ITAIP. Accesible en el sitio: <http://inicio.ifai.org.mx/Publicaciones/La%20promesa%20del%20Gobierno%20Abierto.pdf#page=19>

Belbis, J. I. (2014). Del gobierno abierto a los datos abiertos. En M. Riorda, & L. Elizalde, Comunicación Gubernamental 360. Buenos Aires: La crujía.

Ramírez Alujas, A. V. (2012). Gobierno abierto es la respuesta ¿cuál era la pregunta? Más poder local, 14-22. Disponible en el sitio: [https://www.academia.edu/10034670/Gobierno\\_abierto\\_es\\_la\\_respuesta\\_Cu%C3%A1l\\_era\\_la\\_pregunta](https://www.academia.edu/10034670/Gobierno_abierto_es_la_respuesta_Cu%C3%A1l_era_la_pregunta)

Vallefín, C. A.; López, J. I. (2017) Anotaciones a la ley 27.275 de Acceso a la Información Pública: sinopsis y comparación con el régimen anterior. Thomson Reuters.

Vientos de cambio II: Avances y desafíos de las políticas de gobierno abierto en América Latina y el Caribe - Ramírez Alujas, Álvaro V.; Dassen, Nicolás: <https://publications.iadb.org/es/publicacion/17111/vientos-de-cambio-ii-avances-y-desafios-de-las-politicas-de-gobierno-abierto-en>

Sección de Datos Abiertos - Blog Abierto al Público: <https://blogs.iadb.org/conocimiento-abierto/es/datos-abiertos/>

Scrollini, F. (2018) Automatizar con cautela: Datos e Inteligencia Artificial en América Latina. ILDA. Disponible en el sitio <https://idatosabiertos.org/automatizar-con-cautela-datos-e-inteligencia-artificial-en-america-latina/>

## MÓDULO 2

---

# Publicando datos abiertos desde el sector público

---

Docente: Dr. Juan Pane

TUTORIAS

Susana Soto  
Daniel Villatoro

# Resumen introdutorio

El sector público gestiona una gran cantidad de datos sobre los servicios provistos a la ciudadanía y la infraestructura del país. La publicación de estos datos puede habilitar la colaboración en la innovación.

En este módulo se introducirá a los servidores públicos en estrategias y herramientas para la liberación de datos en el sector público. Los objetivos de este módulo se enfocan en analizar la información presentada y en aplicar los elementos conceptuales aprendidos tanto en este módulo como en el anterior. En particular este módulo cubrirá:

- Desarrollo de una estrategia de liberación de Datos Abiertos: Qué liberar, cuándo liberar y cómo liberar. Introducción a las formas, estrategias y modelos para la liberación de datos para que estos sean efectivos y eficientes.
- Priorización de datos a liberar: indicadores y herramientas (Open Data Readiness Assessment, Open Data Index, Open Data Barometer). Introducción a los índices y herramientas que permiten medir la calidad de los datos abiertos para presentar buenas prácticas que pueden ser útiles para abrir datos.
- Arquitecturas y alternativas de plataformas para liberar datos: portales de datos abiertos, introducción a sistemas de bases de datos (CKAN - Github) y a repositorios de gestión documental (Eprints - DSpace).
- Estándares para liberación de datos e instrumentos de medición: Introducción a los estándares y modelos de calidad de datos abiertos, y la Carta Internacional de Datos Abiertos; análisis de sus principales características y ejemplos de países que la han adoptado en la región. Introducción a las principales mediciones internacionales en el área de datos abiertos. Principales características, fortalezas y debilidades de cada una de ellas.
- Contrataciones Abiertas: Introducción al modelo de Open Contracting (OCDS) y sistemas como Popolo.

## OBJETIVOS DE APRENDIZAJE

1. Comprender los conceptos clave relativos a la apertura de datos
2. Contrastar las diferentes estrategias y alternativas de publicación de datos
3. Comprender los indicadores y estándares de datos abiertos
4. Comprender los conceptos básicos de las Contrataciones Abiertas y el Estándar de Datos para las Contrataciones Abiertas

## GLOSARIO DE TÉRMINOS

**Dato:** es una representación simbólica que describe hechos, condiciones, valores o situaciones

**Información:** es un conjunto organizado de datos procesados y relacionados de manera que nos permitan comunicar o adquirir conocimientos

**Datos abiertos:** son datos que pueden ser libremente usados, reusados y redistribuidos por cualquiera, sujeto solamente, a lo sumo, a requisitos de atribución y distribución con la misma licencia

**Licencia de uso de datos abiertos:** es la autorización de uso de los datos emitida por la fuente que posee los derechos de autor de los datos.

**Conjunto de datos:** es una colección de registros organizados de datos en donde cada elemento tiene la misma estructura, ordenados para su procesamiento por un ordenador

**Catálogo de datos abiertos:** es un sistema de gestión de información que tiene por objetivo ser un punto de referencia único para quien desee buscar datos y acceder a los mismos. El mismo se conforma por un sistema de gestión de conjuntos de datos (datasets) y sus metadatos, y pone a disposición de los usuarios herramientas para agilizar la publicación, el acceso, la búsqueda, la navegación y la utilización de los datos

**Open Contracting:** Es una organización sin fines de lucro que busca acelerar el ecosistema de transparencia y participación ciudadana relacionada a las compras y contrataciones públicas en el mundo mediante la publicación de dichos datos en formatos de datos abiertos y el involucramiento de los actores relacionados en el ecosistema.

**OCDS:** El Open Contracting Data Standard (o Estándar de Datos para las Contrataciones Abiertas - EDCA) es un estándar basado en json que define qué datos de los procesos de contrataciones públicas se deben publicar, y de qué manera.

## Desarrollo de una estrategia de liberación de Datos Abiertos

### Introducción

En este módulo se aprenderá a desarrollar una estrategia de publicación de datos. La misma debe responder a las siguientes preguntas. Qué liberar, cuándo liberar y cómo liberar datos e información para que los procesos de publicación y el uso de estos sean efectivos y eficientes.

#### ¿Qué liberar?

Datos, contenidos en conjuntos de datos, con sus correspondientes metadatos

#### ¿Cómo liberar?

En conjuntos de datos, que pueden tener varias distribuciones y formatos. Cada publicación de conjunto de datos debe tener al menos los siguientes archivos

- Un archivo con los datos,
- Metadatos (ver sección de metadatos) de:
  - Los datos: por ejemplo, el significado de cada variable
  - El conjunto de datos: por ejemplo, la licencia, periodicidad de actualización de datos, validez temporal.

#### ¿Cuándo liberar?

La definición de cuándo liberar, y quiénes son los responsables del trabajo suele quedar definida en un documento de políticas de publicación. Dicho documento debe ser redactado por las instituciones que gestionan cada uno de los datos, siguiendo las directrices del organismo que gestiona o regula la publicación de datos a nivel nacional, local o provincial.

### 1.1 Diferencia entre Datos e Información

Un dato es una representación simbólica que describe hechos, condiciones, valores o situaciones, recabada y dispuesta de manera adecuada para cumplir el objetivo de la institución que lo gestiona.

Por cuenta propia estos valores carecen de un valor semántico, es decir que no cuentan con un significado para alguien, por lo que no aportan valor al receptor del mensaje. Para que estos datos tengan sentido, deben ser procesados, asociados o agrupados dentro de un mismo contexto para formar una información.

Por lo expuesto anteriormente, podemos llegar a la conclusión de que una información es un conjunto organizado de datos procesados y relacionados de manera que nos permitan comunicar o adquirir conocimientos.

A modo de ejemplificar estos dos conceptos (dato e información), tomemos en cuenta una encuesta en la que se solicitan los siguientes datos de personas:

Nombre: Juan  
Apellido: Perez  
Edad: 22 años  
Salario Mensual: \$1200

De por sí, cada uno de estos datos no provee una información que pueda servir para tomar decisiones, pero una vez asociados, agrupados y procesados se puede obtener la información de que el salario promedio está por debajo de x cantidad de dólares en relación a la canasta básica de alimentos.

### 1.3 ¿Qué son los conjuntos de datos?

Un conjunto de datos es una colección de registros organizados de datos en donde cada elemento tiene la misma estructura, ordenados para su procesamiento por un ordenador. Por ejemplo, un conjunto de datos puede ser el listado de los establecimientos escolares del País, el listado de todos los contratos del Estado para todas sus instituciones, o el Presupuesto General de la Nación.

El mismo conjunto de datos puede tener varias distribuciones (o recursos) que pueden variar en dos dimensiones:

**Temporal:** en este caso el mismo conjunto de datos tiene registros asociados a un tiempo dato. Por ejemplo, el Presupuesto General de la Nación tiene una versión diferente cada año, al igual que la lista de contratos de un Gobierno.

**Formato:** cada conjunto de datos puede ser representado en varios formatos. Por ejemplo, si consideramos que la lista de contratos de un gobierno, puede ser representada en una tabla, la misma puede estar digitalizada para ser abierta con Acrobat Reader (en formato .pdf), o Microsoft Excel (.xls), por cualquier procesador de texto (.csv) o para ser procesado por sistemas automatizados (.json), entre otros.

### 1.4 Importancia de los metadatos

Los Metadatos son datos o descripciones estructuradas que describen las características de otros datos o informaciones. La importancia de los metadatos radica en que proveen contexto para interpretar los datos o las informaciones que acompañan.

A modo de ejemplo podemos tomar en cuenta la temperatura. Si decimos 32 grados no podemos interpretar a cabalidad el significado, o podría malinterpretarse, ya que sin la información adicional de unidad de medida, Celsius o en Fahrenheit, 32 grados puede significar que hace frío (32°F = 0°C) o que hace calor (32°C). Es decir que por medio de este dato (metadato) adicional llegamos a la "información" de que hace frío o calor.

Los metadatos pueden aplicarse a los datos, y a los conjuntos de datos:

**Metadatos de datos:** Estos describen la estructura de los datos y proveen información de cómo interpretarlos. Por ejemplo, la lista de las columnas, las descripciones de los valores de cada columna, el formato (en el caso de fecha si esta se encuentra en dd/mm/aaaa o mm/dd/aaaa) o en el caso de montos la moneda en que están representados los mismos, entre otros.

**Metadatos de conjuntos de datos:** Describen al conjunto de datos en sí. Por ejemplo, qué institución los produjo, la persona de referencia en caso que se necesite realizar consultas, la fecha en que se creó el conjunto de

datos y la validez temporal de los mismos. Normalmente estas definiciones están dadas por las políticas de publicación de la institución que mantiene y publica los datos, y si bien deberían ser publicados en formato de documento en la misma, es también importante publicar la información contenida en las políticas de publicación como metadatos que acompañan a los datos.

Los metadatos son también presentados como los diccionarios de datos.

## 1.5 Guía de buenas prácticas para la Publicación de datos

Las políticas de publicación tienen como objetivo asegurar que la información se encuentre de forma accesible en línea y en formatos abiertos, para que pueda ser usada, reutilizada y redistribuida por cualquier interesado.

A continuación se lista una serie de preguntas que ayudan a guiar el desarrollo de las políticas de publicación para portales de datos abiertos:

¿Qué datos se publicarán iterativamente y cuándo?

Se refiere a la hoja de ruta (roadmap) para publicar la información. Dado que generalmente se cuentan con recursos limitados, es difícil publicar inicialmente el 100% de toda la información disponible, por lo que resulta importante contar con una hoja de ruta a modo de contar con objetivos claros y priorizados con relación a que se va a publicar y cuándo se logrará.

¿Dónde se publicarán los datos? ¿Cómo se publicarán los datos?

Se refiere a la decisión de la dirección web en donde estará el portal de datos abiertos, como así también las decisiones con relación a los formatos en los cuales se publicarán los datos (JSON, CSV, JSON-LD), si incluirá una API para desarrolladores, descargas masivas, buscadores especializados, etc

¿Cuál es la frecuencia de actualización de los datos?

Se debe tener en cuenta que pueden existir conjuntos de datos que necesiten una frecuencia mayor de actualización que otros, pudiéndose por ejemplo indicar algunos como actualizaciones diarias (nocturnas, medio día, etc), semanales, mensuales, etc.

¿Quién es responsable por la publicación de los datos?

Se refiere a quienes son los responsables de la gestión de los datos (sistema, institución, etc.) considerando que:

- Los publicadores de los datos son instituciones que generan los mismos datos, y son responsables por la calidad de los mismos, por ejemplo, los datos de educación publicados por un Ministerio de Educación.
- Los publicadores de los datos son instituciones que proveen plataformas de gestión de datos, que otras instituciones utilizan para gestionar sus datos. Por ejemplo, una Agencia Nacional de Compras pone a servicios de otras instituciones del estado una plataforma de gestión de procesos de compras públicas (comúnmente conocido como e-procurement system), pero en el cual el responsable de la calidad de los datos es cada institución compradora, y no la publicadora de los datos (que es la Agencia Nacional de Compras).

En todos los casos se debe especificar quién publica el dato, y quién es responsable por mantener la veracidad y calidad de los datos.

¿A quién contactar en caso de tener consultas?

Es importante siempre dar a conocer como realizar consultas relacionada a los datos, a fin de evitar malos entendimientos en caso de que alguien necesite más detalles.

¿Qué licencia de publicación de los datos se utilizará?

La licencia define los permisos que otorga el propietario de los datos en relación a lo que los usuarios pueden hacer. Una licencia abierta (para datos abiertos) debe a lo sumo requerir atribución a la fuente y redistribución con la misma licencia.

¿Dónde se puede encontrar mayor información de referencia?

Se debe contar con un lugar dentro del portal de datos abiertos en el que se pueda acceder a más información sobre los temas relacionados, donde se encuentran los diccionarios de datos, manuales de datos, o contar con enlaces a sitios en donde encontrarlos.

¿Cuál es el marco normativo de referencia?

Es muy importante siempre contar con todas las referencias que sean necesarias relacionadas a leyes, decretos, resoluciones, circulares, etc., que sirvan de referencia a todo lo que se encuentre expuesto en el portal de datos abiertos.

## Indicadores internacionales de datos abiertos y esquema de publicación

### 2.1 Introducción

En esta unidad conoceremos el esquema definido por Tim Berners-Lee (creador de la World Wide Web) para calificar cómo se encuentran distribuidos los datos abiertos bajo el concepto de las 5 estrellas. Aprenderemos como cada estrella representa un escalón que nos conduce a datos abiertos más robustos y con mayores funcionalidades y posibilidades. También hablaremos sobre los tipos de datos y los formatos de archivos, entendiendo la relación que debería existir entre estos dos para ser publicados.

Esta unidad nos presenta también el concepto de indicadores de control para los portales de datos abiertos, enseñándonos para qué sirven estos indicadores y cómo nos pueden ayudar a medir la situación de nuestros portales de datos abiertos analizando dos grandes indicadores: Open Data Index y Open Data Barometer.

Desde el punto de vista de una política nacional de datos abiertos gubernamentales, es importante conocer los indicadores y considerarlos para la creación de la estrategia de publicación, ya que la evaluación internacional de la política de datos abiertos se hará en relación a los conjuntos de datos de los indicadores, y respondiendo a las preguntas definidas en ellos.

En el caso de Uruguay se recomienda poner foco en el usuario/a, consultar la demanda de datos y tomar los resultados para priorizar los datos a publicar. En general no se recomienda tener como base los indicadores de medición sino lo que los usuarios/as necesitan, así como el apoyo a las políticas públicas. La mejora en los indicadores debería ser un efecto secundario de esta priorización.

### 2.2 Datos a 5 estrellas

El enfoque técnico de la apertura de datos propuesto está basado en el esquema de apertura de datos a cinco estrellas definido por Tim Berners-Lee, un resumen del mismo se puede observar en la Figura 1. Esta sección presenta el contenido traducido y adaptado del estándar definido en <http://5stardata.info/>.

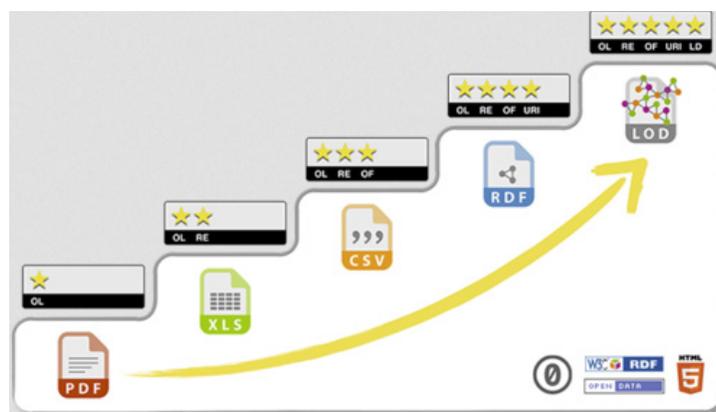


Figura 1  
Esquema de Datos a 5  
estrellas por Tim Berners-Lee  
<http://5stardata.info/>

Este esquema propone una escala incremental de niveles de apertura de datos, donde cada nivel supone avances en términos de los objetivos de los datos abiertos: libertad de uso, re-uso y redistribución de los datos.

Para una explicación narrada de esta sección, ver el video de Presentación de Open Government Data desde 33:13 en adelante (<https://www.youtube.com/watch?v=NiS4i0XKmVE> )

Para leer una discusión sobre este tema se puede ver más en <https://es.schoolofdata.org/2017/07/12/la-locura-por-la-quinta-estrella/>

## 2.3 El gran salto a la tercera estrella

La tercera estrella implica que los datos se encuentren en un formato no propietario, es decir que los datos pueden ser consumidos por cualquier persona y esta puede volver a procesarlos. Por este motivo, la comunidad se encuentra trabajando para estandarizar los formatos abiertos a ser utilizados, con el objetivo de facilitar el trabajo de los consumidores de datos. Estos formatos se resumen en la siguiente tabla:

Formato	Descripción	Esquema Estándar Propuesto
<a href="#">CSV</a>	Formato tabular de datos donde las columnas se delimitan con una coma, aunque comúnmente se aceptan otros separadores como el punto y coma. Aunque no se encuentra estandarizado, existen esfuerzos por definir buenas prácticas como el RFC 4810.	<a href="#">JSON Table Schema</a>
<a href="#">JSON</a>	Formato de intercambio de datos basado en el esquema clave-valor, inspirado en el modelo de objetos de Javascript. La principal diferencia con el formato CSV es la posibilidad de definir estructuras anidadas.	<a href="#">JSON Schema</a>

## 2.4 Tipos de Datos vs. Formatos

Un punto importante a tener en cuenta es que dependiendo de los tipos de datos que se quieren publicar, existen diferentes formatos a ser utilizados. Por ejemplo, si los datos son tabulares, es decir que podemos verlos como datos contenidos en una tabla, uno de los formatos más utilizados es el CSV, en cambio, si los datos indican geo-referenciamiento existen otros formatos especializados en representar esta información.

A continuación se presentan algunos tipos de datos y los formatos más comúnmente utilizados:

Tipos de datos	Descripción	Formatos comunes
<b>Datos genéricos</b>	Datos que no tienen aplicaciones especializadas, y que normalmente se corresponden con datos provenientes de bases de datos y reportes como planillas de cálculo y cuadros de información	Las características de los formatos a usar son: 1. Se debe considerar el uso de formatos estándares, abiertos (no propietarios). En casos muy justificados (datos históricos, transformaciones costosas), usar formatos propietarios. 2. Para datos de planillas y bases de datos tradicionales, se debe considerar el uso de formatos CSV y JSON.
<b>Imágenes</b>	Datos que provienen esencialmente de fotografías	1. Se debe usar tipos de archivos de imágenes basados en el método JPEG, y los archivos PNG. 2. Para formatos vectoriales y de visualización en la Web que necesitan renderización, se recomienda SVG (scalable vector graphics).
<b>Datos estadísticos</b>	Datos usados por usuarios especializados del área estadística. Están en formatos particulares para reúso y explotación de aplicaciones estadísticas especializadas.	Como formato de datos se recomienda el Statistical Data and Metadata Exchange (SDMX) (estándar OCDE). Por su amplio uso son aceptables: SPSS, STATA (en uso hoy día). En este último caso se debe considerar además publicar una distribución de los datos en formato CSV.
<b>Datos georeferenciados</b>	Datos geográficos diseñados para capturar, almacenar, manipular, analizar y desplegar en todas sus formas la información geográficamente referenciada	Entre los formatos más comunes podemos encontrar: .kml, .geojson, .topojson

## 2.5 Indicadores de Control para Datos Publicados

Los indicadores sirven para evaluar el cumplimiento de los principios de datos abiertos que son aplicados en las iniciativas de datos abiertos a nivel internacional. Esto permite que dichas iniciativas sean públicamente auditables por la ciudadanía y comparables entre sí.

Los indicadores internacionales de datos abiertos más comúnmente utilizados, que evalúan el desempeño de un Gobierno en materia de

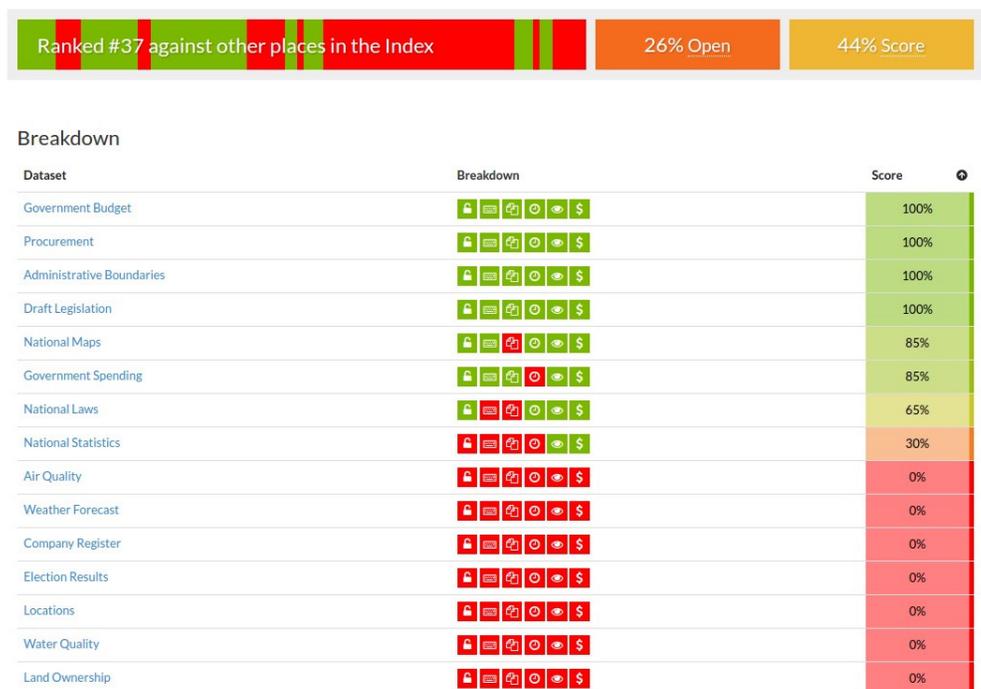
Datos Abiertos, son el Open Data Index y el Open Data Barometer. La diferencia fundamental entre ambos radica en que el Open Data Index evalúa solamente la disponibilidad de los datos, sin embargo, el Open Data Barometer, además de evaluar la disponibilidad, evalúa la utilidad de los mismos mediante una serie de instrumentos como el uso de los datos, el impacto político, económico y social y otros indicadores.

El Open Data index contiene una referencia anual de la situación con relación a cómo los gobiernos publican sus datos midiendo la apertura de sus datos de acuerdo a la definición de Conocimiento Abierto <http://opendefinition.org/od/2.1/es/>. Este índice ha sido compilado en los años 2013, 2014, 2015 y 2016. En este sentido, en Open Data Index es una fuente de referencia sobre la evolución de los distintos países en relación a la publicación de sus datos abiertos gubernamentales que puede servir de información sobre cómo otros países han creado y evolucionado en el desarrollo de sus ecosistemas de datos abiertos.

Las condiciones evaluadas por el Open Data Index son las siguientes:

1. ¿Existe el dato?
2. ¿Está digitalizado?
3. ¿Está públicamente disponible?
4. ¿Es gratuito?
5. ¿Está publicado en Internet?
6. ¿Es procesable por máquina?
7. ¿Se puede descargar en forma masiva?
8. ¿Tiene una licencia abierta?
9. ¿Los datos se proporcionan en tiempo y actualizados?

Pantalla de Evaluación de Datos Abiertos del Open Data Index



El Open Data Barometer tiene como objetivo descubrir el impacto de las iniciativas de datos abiertos en todo el mundo. Analiza las tendencias mundiales y proporciona datos comparativos sobre países y regiones utilizando una metodología en profundidad que combina datos contextuales, evaluaciones técnicas e indicadores secundarios. El Open Data Barometer, a inicios del 2019, cuenta con cuatro ediciones publicadas: 2013, 2014, 2015, 2016. Más información puede ser encontrada en su sitio oficial.

Las preguntas que el Open Data Barometer utiliza para evaluar los conjuntos de datos publicados por los gobiernos son:

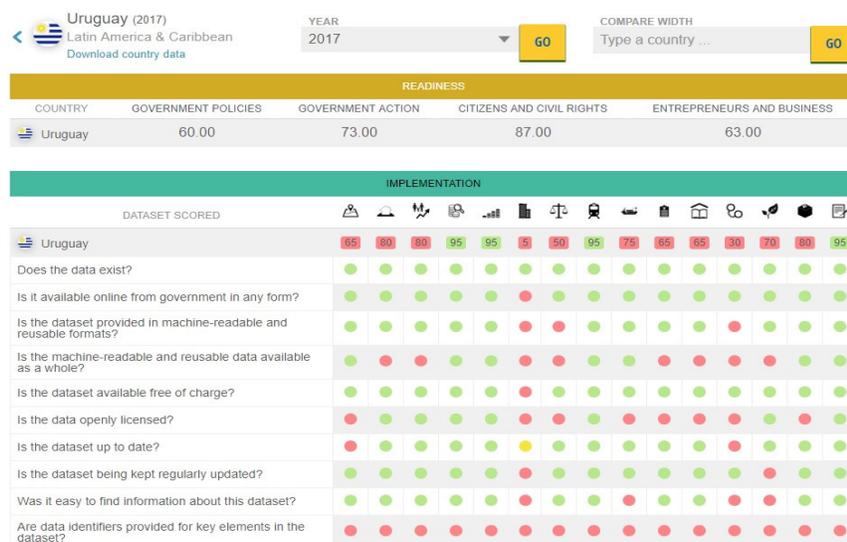
- ¿Existe el dato?
- ¿Está disponible en línea desde el gobierno en cualquier forma?
- ¿Es disponible el conjunto de datos en formato procesable por máquinas?
- ¿Pueden descargarse en forma masiva los datos procesable por máquinas?
- ¿Está el conjunto de datos disponible de forma gratuita?
- ¿Tienen los datos una licencia abierta?
- ¿Está actualizado el conjunto de datos?
- ¿Se publica el dato en forma sostenible?
- ¿Fue fácil encontrar información sobre este conjunto de datos?
- ¿Se proporcionan URIs (identificadores únicos desreferenciables o identificadores de recursos uniformes en Internet) para los elementos claves del conjunto de datos?

Además de evaluar la calidad de los datos en las distintas áreas, el Open Data Barometer evalúa además (y a diferencia del Open Data Index):

El contexto y la preparación (readiness) del ecosistema de datos abiertos gubernamentales (open data context). Por ejemplo, si existen recursos presupuestarios destinados a la iniciativa.

El impacto de los datos publicados en el periodo evaluado (open data impact), por ejemplo, mediante menciones en los medios de prensa.

En la siguiente figura se puede observar el resultado del Open Data Barometer para Uruguay para el año 2017



Pantalla de Evaluación de un país en el Open Data Barómetro

Cualquier iniciativa de datos abiertos que publique datos debe tener en cuenta estos principios y criterios de evaluación para mejorar la calidad de sus datos, su usabilidad, y su potencial impacto.

# Arquitecturas y alternativas de plataformas para liberar datos

### 3.1 Introducción

En esta unidad entraremos en detalles sobre qué tener en cuenta para crear un portal de datos abiertos, desde los conceptos básicos hasta los posibles usuarios finales. Hablaremos sobre la forma de exponer los datos, ya sean por medio de archivos de descarga masiva (bulk download) o interfaces programáticas de aplicaciones (APIs) dirigidas a desarrolladores de sistemas informáticos. Entraremos también en detalles sobre catalogación de datos para luego hablar sobre qué es una arquitectura de datos abiertos y qué cosas tener en cuenta para definirla.

Finalmente veremos que en algunas situaciones existe necesidad de publicar documentos, para la cual existen protocolos de intercambio de metadatos de documentos y herramientas que implementan dichos protocolos, facilitando así la integración de los mismos.

### 3.2 ¿Qué es un catálogo de datos abiertos?

El catálogo central de datos tiene por objetivo ser un punto de referencia único para quien desee buscar datos y acceder a los mismos. El mismo se conforma por un sistema de gestión de conjuntos de datos (datasets) y sus metadatos, y pone a disposición de los usuarios herramientas para agilizar la publicación, el acceso, la búsqueda, la navegación y la utilización de los datos.

Existen dos tipos de usuarios del catálogo:

**Productores de información:** Los productores de información son los que generan los conjuntos de datos con sus metadatos, y desean que los mismos sean catalogados, así como utilizados por los consumidores de información.

**Consumidores de información:** Son los usuarios que desean acceder a la información y los datos que son publicados. Generalmente se pueden dividir a los consumidores en dos grupos:

**Personas:** estos usuarios desean navegar y visualizar los datos para estudiarlos. Este tipo de usuario necesita interfaces gráficas de usuario (GUI, del inglés Graphical User Interface) intuitivas y públicas en HTML.

**Máquinas:** estos usuarios acceden a la información catalogada de manera automática mediante una interfaz de programación de aplicación (API, del inglés Application Programming Interface) bien definida de manera a integrar la misma con otras aplicaciones. Se puede considerar que el Portal de Datos Abiertos reside en esta categoría.

Normalmente el catálogo de datos abiertos posee un índice que permite que las búsquedas sean rápidas y eficientes. En índice puede opcionalmente también permitir la búsqueda de los conjuntos de datos no solo por sus metadatos (nombre, publicador, formatos, etc) sino también por el contenido de los conjuntos de datos.

El catálogo puede también, opcionalmente, ofrecer la capacidad de almacenar localmente los datos. Este servicio puede ser útil en los casos que los proveedores de los datos presenten problemas para publicar sus datos, por ejemplo, por falta de infraestructura tecnológica como servidores o conexión a Internet.

Existen distintas herramientas de catálogos de datos abiertos como ser, entre otros,

CKAN (<http://ckan.org/>) es un software gratuito de código fuente abierto desarrollado por la Open Knowledge Foundation. CKAN permite la gestión de los conjuntos de datos y busca hacer que los mismos sean los más accesibles y fáciles de usar. Como el proyecto es open source, existe una gran comunidad de desarrolladores a nivel mundial que han creado extensiones de CKAN para adaptarlo a sus necesidades de publicación de datos, o bien a sus necesidades de estilos visuales. CKAN es usado por varios gobiernos a nivel nacional y subnacional, entre los que se puede destacar Australia, Brasil, Argentina, Uruguay, USA, Canadá, Berlín, Suiza, EU, Paraguay, entre muchos otros.

Socrata (<https://socrata.com/>), es un sistema cerrado de gestión de datos como servicio en la nube centrado especialmente en gobiernos. El sistema es desarrollado por Tyler Technologies. Si bien el software no es de código fuente abierto, provee una Interfaz Programática de Aplicación (API: Application Programming Interface) que permite que otros desarrolladores puedan crear sistemas y servicios integrados a Socrata. Socrata provee una interfaz de usuario fácil de usar y funcionalidades para crear visualizaciones en base a los datos almacenados en el mismo.

Junar (<http://www.junar.com>): Junar es también una plataforma para la gestión de datos como servicio. Así como Socrata, no es de código fuente abierto y también provee una API para el desarrollo de aplicaciones en base a los datos almacenados en el mismo. A diferencia de Socrata, Junar disponía de una versión gratuita, con funcionalidades reducidas, denominada Datal. El repositorio del proyecto se encontraba anteriormente disponible en <https://github.com/datal-org>.

La selección de los procesos de publicación de datos y de catalogación de los mismos dependerá de los recursos computacionales, monetarios y el equipo técnico disponible para dedicar a la tarea de publicación de datos en formato de datos abiertos.

### **3.3 Portal de Datos Abiertos**

El portal para datos abiertos es el sitio que contiene toda la información referente al ecosistema datos abiertos, al proceso de apertura de datos, y al uso de los datos. El objetivo del portal es la creación de una comunidad interesada en el uso, re-uso y redistribución de datos abiertos, proporcionando todas las herramientas necesarias como:

- Materiales de capacitación de datos abiertos.
- Estándares de datos abiertos.
- Manuales de referencia para las diferentes herramientas de la arquitectura de datos abiertos.
- Aplicaciones (web y móviles) basadas en datos abiertos.

- Foros de consulta sobre temas relacionados a datos abiertos.
- Herramientas de solicitud de apertura de datos, ya sea para la priorización de datos a ser publicados como solicitudes formales de acceso a la información pública esperando como respuesta datos en formato de datos abiertos.
- Índice de eventos relacionados a datos abiertos, como pueden ser hackatones y reuniones de capacitación, de socialización y de intercambio de experiencias.
- Noticias relacionadas con los datos abiertos a nivel local y mundial.
- Lista de contactos de expertos en las diferentes áreas relacionadas a datos abiertos.
- Estadísticas de publicación y uso de datos abiertos.
- Listado de conjuntos de datos
- Diccionarios de datos
- Buscadores de información

Como se puede notar, el portal de datos abiertos es más extenso que el catálogo de datos abiertos, ya que el catálogo es solamente una de las funcionalidades del portal. El catálogo puede estar integrado al portal (ejemplo: DKAN), como también puede estar disponible de manera separada (ejemplo: CKAN).

### 3.4 El Portal de Datos Abiertos y sus Tipos de Usuarios

Un portal de datos abiertos puede contar con diferentes tipos de audiencias o usuarios (personas), estos pueden ser:

**Público en general:** quienes deberían poder acceder a vistas simplificadas que muestren todos los conjuntos de datos y permitan hacer búsquedas sobre ellos. Por ejemplo, para esta audiencia sería útil la representación de los datos en tablas que permitan observar los datos, filtrarlos por columna, ordenar columnas, etc. Sería muy parecido a contar con los datos y abrirlos con un editor de hojas de cálculo como Microsoft Excel o Libre Office. Ejemplo: <https://contrataciones.gov.py/datos/data/contratos>

**Analistas de datos:** este tipo de audiencia no solo tiene curiosidad por ver los datos en bruto sino que quiere poder realizar análisis sobre ellos, buscando extraer información para comprender mejor algún fenómeno o problema, y buscar una posible solución. Este tipo de audiencia incluye, entre otros, a economistas, periodistas, investigadores, etc. Para este tipo de audiencia se puede proveer información graficada (visualizaciones), que permita notar con rapidez las diferencias importantes entre los datos, que si bien podrían ser encontradas también en representaciones por medio de tablas como comentábamos arriba, no serían tan fáciles de notar las diferencias.

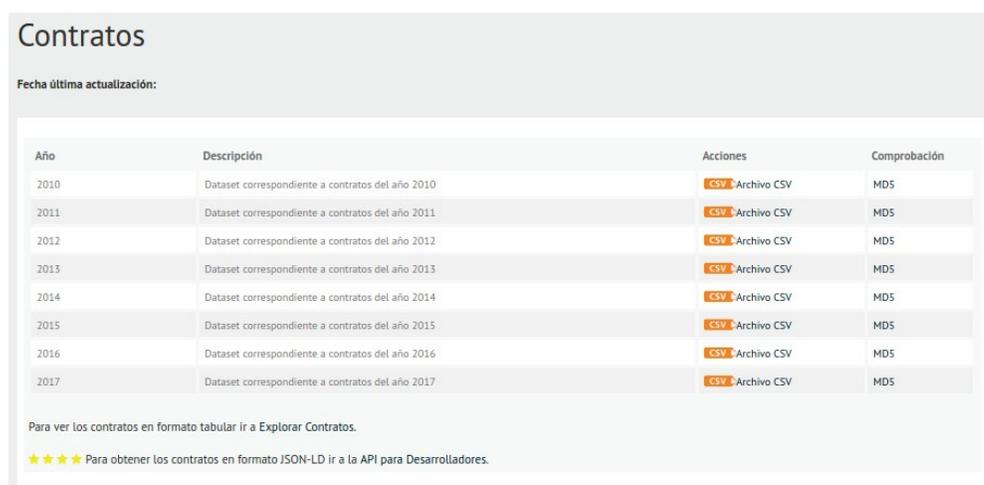
**Informáticos/Desarrolladores:** este tipo de audiencia permite extender las funcionalidades de un portal de datos abiertos ya que poseen conocimientos para usar los datos, procesarlos, y crear nuevos servicios y aplicaciones que permitan generar valor agregado e innovar en base a las necesidades ciudadanas.

Considerando que el Estado cuenta con recursos limitados para desarrollar todos los productos y servicios informáticos que desea, la comunidad de desarrolladores puede dar una gran ayuda en la creación de los servicios

necesarios. Por ejemplo, si los datos de las contrataciones públicas fueran publicados mediante interfaces programáticas de aplicación (APIs), aún mejor, siguiendo estándares internacionales, cualquier programador podría tomar esos datos expuestos y crear una aplicación móvil que permita a potenciales oferentes enterarse rápidamente sobre nuevas licitaciones del Estado en los rubros que le son de interés. Ejemplo: <https://www.contrataciones.gov.py/datos/api/v2/>

### 3.5 Descarga masiva de datos

Como vimos en el apartado anterior, las APIs permiten la exposición de los datos para que estos sean reusados, pero además de estas APIs, se debe prever una funcionalidad que permita la descarga de los conjuntos de datos a fin de poder realizar procesamientos masivos más rápidos, sin tener la necesidad de obtenerlos mediante el API. Este concepto es conocido como bulk download, o descargas masivas de datos. La siguiente imagen muestra un ejemplo de página de descarga masiva de contratos, en donde se encuentran archivos de descarga, separados por años.



Contratos

Fecha última actualización:

Año	Descripción	Acciones	Comprobación
2010	Dataset correspondiente a contratos del año 2010	CSV Archivo CSV	MD5
2011	Dataset correspondiente a contratos del año 2011	CSV Archivo CSV	MD5
2012	Dataset correspondiente a contratos del año 2012	CSV Archivo CSV	MD5
2013	Dataset correspondiente a contratos del año 2013	CSV Archivo CSV	MD5
2014	Dataset correspondiente a contratos del año 2014	CSV Archivo CSV	MD5
2015	Dataset correspondiente a contratos del año 2015	CSV Archivo CSV	MD5
2016	Dataset correspondiente a contratos del año 2016	CSV Archivo CSV	MD5
2017	Dataset correspondiente a contratos del año 2017	CSV Archivo CSV	MD5

Para ver los contratos en formato tabular ir a [Explorar Contratos](#).

★★★★ Para obtener los contratos en formato JSON-LD ir a [la API para Desarrolladores](#).

Interfaz de Descarga Masiva de Datos de Contratos con el Estado Paraguayo

<https://www.contrataciones.gov.py/datos/contratos>

### 3.6 Arquitectura del Portal de datos abiertos

La arquitectura del portal de datos abiertos es el diseño en el cual se define como se harán las conexiones entre los sistemas informáticos que contienen los datos en bruto de cada una de las instituciones y los sistemas informáticos que efectivamente publicarán los datos. Este diseño, de índole técnico/informático, es la base del portal de datos abiertos y debe ser pensado de acuerdo a las posibilidades técnicas que tenga una institución.

Para el ejemplo tomaremos el caso de un Ministerio de Hacienda, que aunque puede variar entre país y país, tiene generalmente un manejo de los mismos conceptos.

Supongamos que se requiere un portal de datos abiertos en donde sean publicados el presupuesto general de la nación, las licitaciones con el Estado y sus contratos, y los pagos a proveedores del Estado. Estos datos podrían estar en diferentes sistemas informáticos internos del Ministerio. La base de datos que contiene el presupuesto general de la nación podría ser una diferente a la base de datos de las compras del Estado, y a su vez,

estas dos podrían ser diferentes a un sistema independiente en donde se registran los pagos a los proveedores del Estado.

Estos sistemas informáticos, de uso diario para la institución, podrían contener estructuras de datos totalmente diferentes e incluso estar desarrolladas utilizando tecnologías diferentes. El objetivo es centralizar toda la información, de manera integrada, en un portal de datos abiertos para exponerlos todos bajo un estándar de datos abiertos definido por la institución (en la Unidad 5 veremos los estándares técnicos para publicación de datos de transparencia fiscal y contrataciones públicas). Las decisiones que se tomen para lograr esto serán el diseño de la arquitectura del portal de datos abiertos.

Una vez que tengamos tomadas las decisiones de cómo obtener datos de diferentes orígenes de datos, debemos trabajar en tomar las decisiones pertinentes a cómo los vamos a exponer, es decir:

- Cómo vamos a organizar el portal con sus páginas (ver sección: Portal de Datos Abiertos en este mismo módulo),
- El número de estrellas al que queremos llegar y cómo lo lograremos (ver Unidad 2),
- Crearemos un API para exponer los datos a terceros o no,
- Decisiones de seguridad para acceso al API (Autenticación, throttling, etc),
- Centralización de datos en formato estándar en una base de datos que centralice los datos a exponer o obtención y traducción de los datos directamente desde los orígenes de datos, etc.

### 3.7 Repositorios de gestión documental

En algunas ocasiones los “datos” que se requieren almacenar son documentos enteros. Por ejemplo, cuando tenemos el listado de las tesis o trabajos científicos de las universidades, o los documentos de resoluciones de una institución.

Con el fin de facilitar la recolección e integración de los metadatos de los documentos publicados, la iniciativa “Open Archives Initiative” creó un protocolo de recolección de metadatos (OAI-PMH por sus siglas en inglés de Protocol for Metadata Harvesting). De esta manera, si un publicador de información de documentos implementa este protocolo, se pueden crear servicios que consulten las publicaciones y centralicen la información de los documentos. Este protocolo (entre muchos otros) es usado por Google Scholar para indexar los artículos científicos de las universidades que publican sus investigaciones basados en este protocolo.

La herramienta EPrints es gratuita y de código fuente abierto e implementa el protocolo OAI-PMH. Posee muchas de las funcionalidades de sistemas de gestión documentales, facilitando la gestión de documentos (archivado, indexación, búsqueda, visualización, etc.). Sin embargo, EPrints es primordialmente usado para repositorios documentales de instituciones académicas, y en especial para revistas científicas.

DSpace es otra herramienta gratuita y de código fuente abierto para la gestión de colecciones de documentos digitales. Así como EPrints, DSpace también es compatible con el protocolo OAI-PMH y es comúnmente usado como repositorio bibliográfico institucional para alojar libros, tesis, datos de investigación.

La iniciativa ROAR (Registry of Open Access Repositories) mantiene un listado de repositorios de acceso abierto en el mundo.

# Estándares para liberación de datos e instrumentos de medición

## 4.1 Introducción

Los principios de las iniciativas como el Open Data Charter sirven como guías para definir y desarrollar políticas de datos abiertos a alto nivel. En términos prácticos, al momento de responder las preguntas de los componentes de una política de publicación, se debería consultar los principios para definir qué conjuntos de datos priorizar y cuándo publicarlos.

Una vez definidos el qué y cuándo, es importante definir el cómo se van a publicar los datos. Es en este momento que los estándares a distintos niveles son de gran relevancia. El esquema de datos a 5 estrellas definido por Tim Berners-Lee (ver Unidad 2) sirve para definir los formatos de datos a ser usados (el nivel de reusabilidad de los datos), que tienen una relación directamente proporcional al nivel de esfuerzo necesario para la publicación de los datos, esto es, a mayor nivel de estrellas, mayor es el esfuerzo que se requerirá para publicar los datos. Una vez definido el esfuerzo, podemos concentrarnos en los estándares de dominio específicos como el Open Contracting Data Standard que ayudará a definir cómo publicar la información de contrataciones públicas, y con qué formato específico, considerando además los metadatos necesarios (estos estándares serán presentados en la Unidad 5).

URUGUAY

## 3.2 Infraestructura de Datos Espaciales

En octubre de 2013, se crea por los artículos 35 y 36 de la Ley 19.149 la Infraestructura de Datos Espaciales como un órgano desconcentrado de la Presidencia de la República. La Ley fue reglamentada por el Decreto N° 390/014.

Ley N° 19.149 Art. 35 de fecha 11/11/2013 Creación de la Infraestructura de Datos Espaciales

La IDE tiene como cometidos liderar la articulación y el fortalecimiento de la producción y acceso a la información geográfica del Uruguay para que sea fiable, oportuna, interoperable, de alta calidad, y brinde apoyo en el análisis y la toma de decisiones de organismos, academia, empresas y ciudadanos.

Cometidos sustantivos:

- Coordinar, planificar y promover la producción de información geográfica del territorio nacional.
- Garantizar, a través del dictado de normas, estándares y recomendaciones, la interoperabilidad, actualización, calidad y acceso de la información geográfica nacional.
- Integrar la información geográfica producida y utilizada en los diferentes niveles de la Administración, inspirándose en los principios básicos de cooperación y coordinación, así como la transparencia y acceso a la información pública.
- Publicar la información geográfica básica en un GEOPORTAL.
- Favorecer la eficiencia en el gasto público destinado a los distintos

aspectos de la información y sistemas de información geográfica.

- Para el cumplimiento de sus cometidos, podrá comunicarse directamente con toda la administración pública estatal, organismos públicos y entidades privadas.

La IDE ofrece entre otros servicios:

- Servicio unificado para búsquedas direcciones, localidades y sugerencias de calles de Uruguay
- Visualizador <http://ide.uy/visor/>

## ARGENTINA

### **API del Servicio de Normalización de Datos Geográficos de Argentina**

Uno de los servicios desarrollados por la Secretaría de Modernización de la Nación tiene como objetivo el de normalizar los datos geográficos de Argentina. Este caso es destacado por su utilidad en el proceso de estandarización de datos geográficos, permitiendo a cualquier desarrollador consumir los servicios durante el proceso de publicación de sus datos, utilizando un estándar nacional para la descripción de direcciones, localidades y otros puntos geográficos de Argentina.

### Introducción

Los estándares de datos permiten asegurar la interoperabilidad de información. Esto es especialmente importante cuando diversos conjuntos de datos son publicados por diferentes instituciones que pueden contar con una parte de la información sobre un tema en particular. Los estándares proveen además documentación muy importante sobre cómo estructurar los datos y el significado de cada valor, esto es, los diccionarios de datos son estandarizados y compartidos. Con esto, el reuso y la integración de los datos se vuelve más sencilla.

En esta sección se presentan tres estándares de publicación de datos que son necesarios para hacer un seguimiento y control del gasto público:

- Open Contracting Data Standard (OCDS): datos de procesos de contratación e implementación de los contratos, incluyendo los pagos.
- Open Fiscal Data Package (OFDP): datos fiscales de proyectos, sus objetivos, presupuesto general de la nación de ingreso y gastos, detalles de líneas presupuestarias, detalles de procesos de pagos, detalles de gastos que no se refieren a contrataciones públicas.
- Popolo: datos de organizaciones públicas y privadas, personas y sus relaciones con las organizaciones, votaciones, discursos.

### 5.1 Contrataciones Abiertas

Las contrataciones abiertas, según la Open Contracting Partnership (OCP), se refieren a la creación de un ecosistema para que el dinero del gasto público en bienes, servicios y obra pública sea usado de manera más eficiente mediante:

- El trabajo con todos los actores del ecosistema (stakeholders) ya sean los compradores, proveedores del estado, beneficiarios del gasto público, organizaciones de sociedad civil de actual de contralores del gasto, medios de prensa, infomediarios, etc.
- Provisión de información de las contrataciones del estado en formato de datos abiertos.

Para facilitar la provisión y el uso de datos de contrataciones públicas, la Open Contracting Partnership promueve el uso del estándar de datos para las contrataciones abiertas y lo mantienen, con el fin de actualizarlo a las necesidades de los publicadores y usuarios de la información.

### 5.2 Estándar de Datos para las Contrataciones Abiertas

Los estándares de datos permiten asegurar la interoperabilidad de información, en especial si la información del proceso de la contratación del estado se encuentra de manera distribuida en distintos sistemas de gestión de la información mantenidos y operados por distintos actores.

El Estándar de Datos para las Contrataciones Abiertas (EDCA, o en inglés Open Contracting Data Standard - OCDS) define:

- Qué datos del proceso de la contratación publicar y el significado de cada valor, esto es, los diccionarios de datos son estandarizados y compartidos
- Cómo publicar los datos, es decir, el formato del archivo (.json) y su sintaxis (definido por el json schema).

Estas definiciones facilitan la interoperabilidad y comparación de la información, inclusive a nivel de varios países. Como ejemplo podríamos tomar en cuenta las compras del Estado. Un país podría nombrar a la etapa en donde una entidad del Estado espera las propuestas de los oferentes como “licitación”, otro lo podría nombrar como “llamado”, mientras que otro podría llamar a esta etapa como “convocatoria a oferentes”. La función del OCDS es unificar esta nomenclatura y permitir que muchos publicadores la adopten, facilitando así el uso de los datos.

En el OCDS, todas las fases del proceso de compra pública tienen nombres bien definidos (en Inglés: planning, tender, award, contract, implementation; en español: planificación, licitación, adjudicación, contrato, implementación), y se definen además los campos para cada una de las fases. Mediante el uso de nomenclaturas del estándar podríamos llegar a conversar con mucha mayor rapidez a la hora de analizar y vincular informaciones de compras públicas a varios niveles del estado (nacional y sub-nacional) e inclusive entre instituciones del mismo nivel del gobierno. Por ejemplo, es común que la información de las fases un proceso de compra pública esté distribuido en al menos dos instituciones: i) la encargada de seguir el proceso de contratación, incluyendo licitación, adjudicación y contrato, y ii) la que se encarga del aspecto financiero del proceso de contratación, incluyendo planificación (con el correspondiente presupuesto) e implementación (con los correspondientes pagos). El OCDS permite la publicación de la información de todas las fases en forma distribuida, por cada una de las instituciones que gestiona su parte de la información.

La siguiente figura muestra las fases del proceso de contratación definidas en OCDS:



Un componente esencial de OCDS es el Open Contracting Identifier (OCID). El OCID permite integrar la información publicada en forma distribuida de las diferentes fases del mismo proceso de contratación, siempre y cuando cada parte de la información gestionada y publicada haga referencia al mismo OCID. En la figura anterior podemos además observar, que el proceso de contratación, según OCDS, inicia con la planificación del mismo, y termina con los datos de la ejecución del contrato, no solamente con la firma del contrato. La información de la implementación del contrato

incluye tanto información de entrega de los bienes, servicios y obras contratadas, además de los pagos realizados por los mismos.

Se puede ver un ejemplo del estándar aplicado las compras públicas de varios países en los siguientes enlaces:

- Colombia: <https://www.colombiacompra.gov.co/transparencia/gestion-documental/datos-abiertos>
- Mexico: <https://datos.gob.mx/busca/organization/contrataciones-abiertas>
- Paraguay: <https://www.contrataciones.gov.py/datos/api/v2/#!/ocds>
- Uruguay:
  - Compras:  
<https://www.comprasestatales.gub.uy/inicio/datos-abiertos/open-contracting/>
  - Adjudicaciones:  
<https://catalogodatos.gub.uy/dataset/datos-de-compras-en-linea>

Además de facilitar la publicación, la integración y el análisis de los datos de las contrataciones públicas, OCDS también fomenta la creación de herramientas de código fuente abierto. En el siguiente enlace se puede encontrar una lista de herramientas relacionadas a las contrataciones abiertas mantenidas por la Open Contracting Partnership: <https://airtable.com/shrzycSNYRcmVOWSZ/tblhHNGcDXuievZ74?blocks=hide>

### 5.3 Transparencia Fiscal

Si bien una parte considerable del gasto público se realiza mediante las contrataciones del estado, estas corresponden solamente a una parte del total de los gastos de los presupuestos generales de los países. La transparencia fiscal se refiere a la publicación de la información de cómo el gobierno organiza y ejecuta los presupuesto de ingresos y de gastos.

La Iniciativa Global para la transparencia Presupuestaria (GIFT por sus siglas en inglés Global Initiative for Fiscal Transparency) establece una serie de principios que debe guiar la transparencia fiscal a nivel internacional, entre los cuales se pueden mencionar:

- Todas las personas tienen derecho a buscar, recibir e impartir información sobre las políticas fiscales...
- Los gobiernos deben publicar objetivos claros y medibles para la política fiscal agregada, hacer periódicamente un balance del progreso realizado y explicar toda desviación respecto de lo planificado.
- Debe presentarse al público información financiera y no financiera de buena calidad acerca de las actividades fiscales pasadas, presentes y proyectadas...
- Los ciudadanos deben tener el derecho de participar directamente en los debates y las deliberaciones públicas sobre la formulación y la ejecución de las políticas fiscales

GIFT promueve el desarrollo y adopción de un estándar para la transparencia fiscal, el Open Fiscal Data Package (OFDP). OFDP es una especificación técnica para la publicación de datos de presupuestos y gastos de los gobiernos. A diferencia de OCDS, OFDP no define qué datos publicar, ni con qué nivel de agregación, lo cual permite flexibilidad de qué información publicar, pero sí define un formato específico con el cual esta información debe ser publicada. En general, una publicación de OFDP contiene:

- Archivos en formato de datos fiscales, normalmente tablas en formato .CSV, que contienen los datos que se desean publicar. La estructura de este archivo con los datos es flexible, siempre y cuando sea un archivo CSV válido.
- Un archivo que contiene el mapeo de los conceptos publicados en el archivos .CSV a conceptos fiscales definidos en la especificación del Open Fiscal Data Package. Este archivo permite describir el archivo de datos fiscales con el fin de extraer su representación lógica, sin necesidad de precondiciones y asunciones externas.

Además, GIFT impulsa la creación de la herramienta de código fuente abierto Open Spending, que permite trabajar con los datos de OFDP. Existe una instancia pública y gratuita de Open Spending que permite a los usuarios trabajar con la herramienta y publicar sus datos. La herramienta facilita el proceso de mapeo de los datos al OFDP en una interfaz intuitiva, y una vez cargados los datos, permite explorar los datos cargados y crear distintos tipos de visualizaciones de forma dinámica.

En el caso ideal, cada publicación de información de los diferentes componentes de la transparencia fiscal deben poder ser referenciados entre ellos. De esta manera, los usuarios de la información podrán responder, con los conjuntos de datos disponibles, a consultas de los siguientes tipos:

- ¿Cuáles son las líneas presupuestarias de un proyecto de inversión, en cada presupuesto general de la nación, por año?
- Por cada línea presupuestaria, ¿cuáles son los contratos que se suscribieron para alcanzar los objetivos del proyecto?
- Por cada contrato,
  - ¿cuáles son los pagos realizados para este contrato?
  - ¿cuáles son las empresas beneficiarias de los contratos?
  - ¿quiénes son los beneficiarios reales de las empresas?
- Por cada pago,
  - ¿Qué líneas presupuestarias se utilizaron para realizar los pagos?
  - ¿cuánto tomó realizar el proceso de pago?
- Por cada beneficiario del pago:
  - ¿Está relacionado con la función pública?
  - Si fue un candidato electo, ¿ha recibido financiación durante su campaña electoral?, en caso afirmativo, ¿de quién?
- Para cada entidad contratante:
  - ¿Quién es su representante legal, o forma parte de la cadena de decisión de cómo se realiza el gasto público?
  - ¿Han recibido estas personas financiamiento político para sus candidaturas?, en caso afirmativo, ¿poseen estos financistas proveedores nuevos contratos con el organismo contratante?

Como se puede observar, es importante poder integrar información de diferentes unidades de análisis de información (presupuesto, contratos, pagos, entidades del estado, empresas, beneficiarios, funcionarios, financiamiento, etc). Si bien hemos analizado dos estándares importantes para contrataciones públicas y transparencia fiscal, queda por analizar cómo publicar otro tipo de información, de manera estandarizada. En su conjunto, estos tres estándares (OCDS, OFDP, Popolo) podrían ser la base de una iniciativa de transparencia fiscal integral.

## 5.4 Otro estándar de datos: Popolo project

El Proyecto Popolo define modelos de datos y formatos de publicación e intercambio de datos para que las organizaciones puedan reutilizarlas y así minimizar el tiempo de definir cómo publicar datos. Los tipos de datos definidos por Popolo son:

- Personas y organizaciones:
  - Persona
  - Organización
  - Afiliación (membership)
  - Puesto (post)
  - Detalles de contacto
- Mociones y votaciones
  - Moción
  - Evento de votación
  - Conteo de votos
  - Voto
- Otras clases
  - Área
  - Evento
  - Discurso

Además, fomenta el desarrollo de componentes de código fuente abierto reutilizables, que implementan las especificaciones. Esto facilita la implementación de proyectos cívicos y gubernamentales.

Como se puede observar, una implementación ideal de la publicación de datos en un ecosistema de datos abiertos debería publicar información utilizando varios estándares, que se complementan entre ellos, por ejemplo:

**Open Contracting Data Standard (OCDS):** datos de procesos de contratación e implementación de los contratos, incluyendo los pagos.

**Open Fiscal Data Package (OFDP):** datos fiscales de proyectos, sus objetivos, presupuesto general de la nación de ingreso y gastos, detalles de líneas presupuestarias, detalles de procesos de pagos, detalles de gastos que no se refieren a contrataciones públicas.

**Popolo:** datos de organizaciones públicas y privadas, personas y sus relaciones con las organizaciones, votaciones, discursos.

## MÓDULO 2

# Lecturas

¿Qué es un dato?, Curso de fundamento de datos de la “Escuela de datos” <https://es.schoolofdata.org/que-es-un-dat/>

El Manual de Open Data, Open Knowledge Foundation  
<http://opendatahandbook.org/guide/es/>

Datos a 5 estrellas por Tim Berners-Lee. <https://5stardata.info/es/>

Guía de implementación de Datos Abiertos de Uruguay <https://www.agesic.gub.uy/innovaportal/v/2478/2/agesic/guia-para-publicacion-de-datos-abiertos-de-gobierno.html>.

Getting your data together: routes towards an OCDS API. Define varias arquitecturas para la publicación de datos de Contrataciones Públicas, pero las arquitecturas pueden ser pensadas como genéricas para cualquier conjunto de datos. <https://www.open-contracting.org/2016/06/30/getting-data-together-routes-towards-ocds-api/?lang=es>

Es la hora de datos más abiertos en América Latina - Blog Abierto al Público: <https://blogs.iadb.org/conocimiento-abierto/es/datos-mas-abiertos-en-america-latina/>

Muente, Arturo; Serale, Florencia Los datos abiertos en América Latina y el Caribe: <http://dx.doi.org/10.18235/0001202>

## MÓDULO 3

---

# ¿Para qué liberar datos? Entendiendo la demanda

---

Docente: Dra. Carla Bonina

TUTORIAS

Catalina Gobantes  
Camila Salazar

# **Resumen** **Introdutorio**

El valor de los datos abiertos se materializa con su uso. Sin embargo, captar la atención y generar innovación por parte del espectro de usuarios, intermediarios y consumidores de datos es un desafío. ¿Cómo incentivar la participación de la demanda de datos abiertos? ¿Cómo generar y hacer crecer a la comunidad de usuarios de datos?

Este módulo tiene por objetivo entender qué tipo de valor se puede crear con los datos abiertos, quiénes son los principales grupos de usuarios de los datos y cómo planear estrategias para fomentar tanto la participación de esos grupos como la generación de nuevas iniciativas con datos abiertos.

El módulo contiene cinco unidades: la primera unidad introduce la generación de valor con tecnología cívica y datos abiertos. La unidad dos presenta un mapeo de la demanda, y aquellos grupos de usuarios claves, el cual expande conceptos ya introducidos en módulos previos. La unidad tres presenta los instrumentos prácticos para la mejora de servicios públicos con datos abiertos.

La unidad cuatro introduce la perspectiva de plataforma de innovación, y aquellas reglas y herramientas claves para generar productos o servicios con datos abiertos con los emprendedores o innovadores externos. La unidad cinco revisa los principales desafíos para cultivar la demanda de datos abiertos. Para completar el módulo, se espera que los participantes analicen y apliquen los contenidos en el contexto de sus organizaciones.

## **OBJETIVOS DE APRENDIZAJE**

1. Comprender y reconocer los distintos beneficios que se pueden obtener con la tecnología cívica y los datos abiertos.
2. Identificar los tipos de usuarios que componen la demanda de datos abiertos.
3. Comprender las estrategias para mejorar los servicios públicos con datos abiertos.
4. Reconocer los modelos para la generación de beneficios económicos con datos abiertos.
5. Comprender y analizar desafíos y recomendaciones relacionados con una estrategia de apertura de datos y su demanda.

## GLOSARIO DE TÉRMINOS

**Análisis FODA:** FODA es la sigla de Fortalezas, Oportunidades, Debilidades y Amenazas. El análisis FODA es una herramienta que analiza el estado de situación de una organización, proyecto o sector basado en características internas (fortalezas y debilidades) y la situación externa (oportunidades y amenazas).

**Co-creación:** aquellos procesos o actividades donde al menos dos actores (por ejemplo, público, privado, gubernamental o cívico) colaboran en la realización de un proyecto para alcanzar determinado resultado.

**Conjunto de datos:** es una colección de registros organizados de datos en donde cada elemento tiene la misma estructura, ordenados para su procesamiento por un ordenador. Por ejemplo, un conjunto de datos puede ser el listado de los establecimientos escolares de un país, el listado de todos los contratos del Estado para todas sus instituciones, o el Presupuesto General de la Nación.

**Datos:** es una representación simbólica que describe hechos, condiciones, valores o situaciones, recabada y dispuesta de manera adecuada para cumplir el objetivo de la institución que lo gestiona.

**Hackatón (o hackathon):** Maratón de ideas, diseño y prototipado de software, que reúne a especialistas, técnicos programadores, diseñadores y emprendedores para trabajar de manera colaborativa. Las jornadas tienden a durar entre 24 y 48 horas y suelen tener desafíos específicos para guiar el evento.

**Plataforma de innovación digital:** plataformas digitales que ofrecen funcionalidades sobre las cuales otras empresas, desarrolladores o innovadores independientes pueden construir productos, servicios o tecnologías complementarias.

### Introducción

El valor de los datos abiertos radica en su uso. Esta unidad tiene como objetivo cubrir los siguientes aspectos: ¿qué tipos de beneficios pueden generar los datos abiertos y la tecnología cívica? ¿Qué ejemplos existen en la región y el mundo? ¿Cómo se generan estos beneficios? ¿Cuál es el rol de la participación y colaboración para la generación de valor?

En un mundo con creciente pobreza, desigualdad, degradación ambiental e injusticia, la innovación digital con un enfoque social aparece como un rayo de esperanza para resolver tales problemas. La tecnología cívica es un tipo de innovación digital social, en el cual las tecnologías digitales se implementan para mejorar la relación entre los ciudadanos y el gobierno con la finalidad de incluir a más personas en la toma de decisiones públicas, el empoderamiento ciudadano, o bien para mejorar el acceso y entrega de servicios públicos. En muchos casos, y especialmente en la región, los datos abiertos suelen ser un insumo para el desarrollo de innovaciones basadas en tecnología cívica.

En parte, la creciente disponibilidad general y asequibilidad de las nuevas tecnologías digitales han abierto un mundo de posibilidades al contribuir al empoderamiento ciudadano y la innovación de los servicios públicos. Hoy en día, encontramos innovadores de tecnología cívica en una variedad de industrias, en el sector privado y el público. Las tecnologías cívicas pueden ser desarrolladas por organizaciones sin fines de lucro, por empresas, o incluso por el propio gobierno. En Latinoamérica, Dymaxion Labs utiliza datos abiertos, algoritmos de imágenes satelitales y técnicas de aprendizaje automático para generar información en tiempo real, para abordar el crecimiento de asentamientos informales, contribuyendo a mejorar la planificación gubernamental para abordar el acceso a servicios públicos críticos como el saneamiento. En Perú, el medio digital independiente Ojo Público abrió una cantidad de datos sin precedentes sobre el sistema de salud peruano en el proyecto Cuidados Intensivos. El proyecto reveló varios conflictos de intereses, lo que obligó a las autoridades a crear nuevos controles y, por lo tanto, a contribuir con la transparencia y la rendición de cuentas en el país.

### 1.1 La generación de valor de los datos abiertos y la tecnología cívica

Los datos abiertos abren nuevas posibilidades para generar valor, tanto en el ámbito económico, como en el social, ambiental y democrático. Existen varios proyectos de datos abiertos que tienen un objetivo cívico o social que a la vez pueden estar generando beneficios económicos. Si bien los tipos de beneficios que pueden generar los datos abiertos no son necesariamente excluyentes, a continuación los presentamos de manera separada por simplicidad.

## 1.2 Beneficios económicos

Del lado económico, a medida que las empresas, los gobiernos y ciudadanos utilizan o reutilizan los datos abiertos, surgen nuevos productos, procesos y oportunidades comerciales. En 2011, un estudio encargado por la Comisión Europea estimó que el valor económico de la apertura y reutilización de la información del sector público era de aproximadamente 40 millones de euros anuales, solamente para la Unión europea. A escala mundial, McKinsey estimó en 2013 que los datos abiertos pueden contribuir a la generación de entre \$3 y \$5 billones de dólares por año en la economía global. Si bien estas cifras siguen siendo especulativas, abren un abanico de posibilidades a la hora de la generación de valor económico. La unidad 4 se dedica a explorar a profundidad el aspecto económico.

Los datos abiertos han visto la generación de una nueva industria de datos en sí misma. Han surgido nuevas empresas que capturan valor recopilando, analizando y compilando datos abiertos de diferentes fuentes u ofreciendo un servicio directo basado en datos abiertos. The Climate Corporation en EE. UU., por ejemplo, ha sido un caso notable de una compañía que construyó su negocio utilizando datos meteorológicos abiertos, para ser luego adquirida por Monsanto en 2009 por Usd \$ 930 millones. En América Latina, un ejemplo de emprendimiento en la industria de datos abiertos es Junar. En términos simples, Junar ofrece una plataforma digital para la publicación, extracción y análisis de datos abiertos.

## 1.3 Beneficios económicos

Igualmente importante, los datos abiertos se consideran particularmente beneficiosos para aumentar la transparencia, luchar contra la corrupción y promover la inclusión social, a pesar de que el valor de estos beneficios es mucho más difícil de evaluar en términos monetarios. Esta perspectiva de valor ha sido particularmente importante en América Latina, donde la falta de transparencia y rendición de cuentas, así como las necesidades sociales, han estado en el centro de la agenda del movimiento de los datos abiertos en la región.

La preocupación por la transparencia y la rendición de cuentas no es, sin embargo, un problema exclusivamente de los países en desarrollo. Las políticas de datos abiertos en países pioneros—por ejemplo, el Reino Unido—fueron en parte un resultado de la falta de confianza en el sistema político y su rendición de cuentas. De hecho, en la Cumbre Anual de la Alianza para el Gobierno Abierto que se realizó en Londres en 2013, la apertura drástica de datos gubernamentales fue uno de los grandes destacados de la reunión.

Para los gobiernos, además de mejorar la transparencia, los datos abiertos pueden también contribuir a la innovación—es decir, a la mejora de servicios públicos o de sus operaciones a través de su modernización. Por ejemplo, en Uruguay, el proyecto A Tu Servicio del Ministerio de Salud en colaboración con la organización de la sociedad civil DATA pone a disposición de los usuarios un conjunto de datos sobre las instituciones de salud para que éstos tomen una decisión informada a la hora de elegir su prestador. Este caso de co-creación—concepto que desarrollaremos en breve—los beneficiarios son a la vez varios: los usuarios del sistema de salud, quienes se benefician al acceder a mejor información acerca de los proveedores de salud pública, el propio gobierno, el cual utiliza nuevas

tecnologías y datos abiertos para la modernización y transparencia de su gestión y a los propios proveedores de salud, quienes a la vez pueden utilizar el portal como un espacio para ofrecer mayor calidad y respuesta a sus usuarios.

Son muchos también los casos en los que los datos abiertos se construyen con base en el trabajo voluntario, la cooperación y el compromiso cívico. Existen iniciativas de datos abiertos de producción colaborativa entre iguales basada en el dominio público, en los cuales los participantes colaboran, contribuyen con su tiempo e intercambian conocimientos y experiencia para lograr un bien social o de interés público.

Una iniciativa pionera en el ámbito de la localización geográfica fue la del Open Street Map, un proyecto colaborativo para crear un mapa mundial editable gratuito. De manera similar a Wikipedia, Open Street Map es elaborado por una comunidad de colaboradores con conocimientos locales que actualizan y mantienen datos sobre rutas, estaciones de tren, cafés y mucho más en todo el mundo, todo en formato abierto. Este proyecto tiene un valor utilitario inherente: cualquiera con un GPS puede descargarlo y usarlo gratis, y ha demostrado ser particularmente útil en áreas remotas. Sin embargo, esta tarea común tiene un valor mayor. Como en otros casos de datos abiertos, Open Street Map se puede usar como plataforma para crear otros servicios basados en datos de mapeo, tales como transporte, ayuda humanitaria o monitoreo.

#### **1.4 Participación, colaboración y co-creación con tecnología cívica y datos abiertos**

Tanto la tecnología cívica como los datos abiertos prometen una diversidad de beneficios en el ámbito económico al igual que en lo social, político y democrático. El potencial de los datos abiertos en la región hoy, y también a nivel global, se ve particularmente facilitado por un mayor acceso y asequibilidad a la Internet, computadoras y telefonía móvil entre otras tecnologías digitales. Asimismo, la generación y apropiación de sus beneficios resultan de la participación, colaboración o bien co-creación de servicios de una diversidad de actores—dos pilares fundamentales en el marco del Gobierno Abierto.

Los conceptos de participación, colaboración y co-creación están, sin dudas, relacionados. La participación pública o cívica, puede llevar a casos de colaboración o co-creación de servicios. Por ejemplo, una sesión abierta entre gobierno, organizaciones de la sociedad civil y ciudadanos puede dar ideas para priorizar una serie de datos a abrir, los cuales en el tiempo, pueden a la vez llegar a iniciativas concretas entre una organización civil y el gobierno. En otros casos, una sesión abierta puede ser un buen insumo para modernizar un servicio en el gobierno, que no necesariamente es co-creado en colaboración con un externo.

En términos simples, entonces, por co-creación se entienden aquellos procesos o actividades donde al menos dos actores (por ejemplo, público, privado, gubernamental o cívico) colaboran en la realización de un proyecto para alcanzar determinado resultado. En otras palabras, en donde se requiere más de una parte para la creación de valor.

## 1.5 La demanda de datos abiertos

La participación, colaboración o co-creación con datos abiertos es de gran importancia a la hora de entender la demanda. Los ejemplos que presentamos en los apartados anteriores, además de ilustrar los distintos tipos de beneficios que pueden generarse a partir de iniciativas con datos abiertos y tecnología cívica, también muestran que el gobierno es solo uno de los actores claves. A la hora de pensar en quiénes conforman las comunidades de usuarios—demanda—de los datos abiertos, en general, encontramos:

- Organizaciones de la sociedad civil;
- Medios de comunicación y periodistas;
- Empresas, desarrolladores y hackers cívicos;
- Universidades y académicos;
- Gobierno;
- Público en general.

La siguiente unidad profundiza en entender los roles, usos y ejemplos de estos grupos de usuarios, y lo que puede aportar cada uno de ellos para hacer crecer el uso de los datos.

# Quiénes piden datos abiertos, quiénes los usan y para qué

## Introducción

¿Quiénes componen la comunidad de usuarios de datos abiertos? ¿Cuáles son sus roles? Esta unidad tiene por objetivo identificar a la comunidad de usuarios de datos abiertos gubernamentales. Para ello, introducimos a los distintos actores que componen la comunidad de datos abiertos y sus potenciales usos. Es importante notar que, dependiendo de cada contexto, es posible que no todos los actores que aquí introducimos estén necesariamente presentes en un determinado ecosistema. El objetivo es presentar una lista lo más representativa posible, que pueda también servir de mapa para una estrategia de creación de comunidad de datos abiertos.

En esta unidad, profundizaremos en aquellos grupos de usuarios típicos a mayor detalle, con la finalidad de identificar a la demanda y desarrollar estrategias para su involucramiento desde una perspectiva de política pública, en lugar de técnica. Este enfoque es complementario al utilizado en el módulo 2, y lo señalado en términos de las estrategias de publicación de datos y sus tipos de usuarios sigue siendo de especial relevancia.

## 2.1 Organizaciones de la sociedad civil

El también llamado “tercer sector” es, en muchos casos, de los primeros actores en vincularse con la agenda de datos abiertos. Como ya anticipamos en los módulos anteriores, la sociedad civil tiene un rol principal en la apertura de datos, porque muchas veces es quien aboga por datos en el marco de esfuerzos por incrementar la transparencia y la rendición de cuentas. En cuanto a la reutilización, la sociedad civil generalmente se ocupa de coordinar o bien impulsan proyectos o iniciativas de datos abiertos con efectos directos en procesos de escrutinio público, participación ciudadana o incluso la modernización de servicios. Asimismo, los organismos de la sociedad civil son claves para incidir en procesos de apertura, ya que, además de reusar datos, son quienes suelen movilizar la agenda de apertura y su priorización en los gobiernos.

En la región, organizaciones como DATA en Uruguay, Fundación Ciudadanía Inteligente en Chile y SocialTIC en México son ejemplos de actores de la sociedad civil que han logrado incidencia en las agendas de datos abiertos en sus contextos, e incluso han desarrollado innovaciones conjuntas—tales como los casos de Por Mi Barrio y ATuServicio en Uruguay.

La Asociación Civil por la Igualdad y la Justicia (ACIJ) es una organización de la sociedad civil que ha participado en diversos procesos con foco puesto en la reutilización de datos abiertos y la difusión de los derechos asociados a estos temas. Es particularmente destacable el trabajo que han realizado junto con Wingu y con el apoyo de ALTEC para el desarrollo de “Caminos de la Villa”, una plataforma que está centrada en promover los procesos de urbanización de las villas ubicadas en la Ciudad de Buenos Aires, incorporando datos públicos y generando nuevos datos que son luego abiertos a través de la plataforma, especialmente sobre la provisión de servicios públicos y los circuitos de tránsito interno.

## 2.2 Organizaciones de la sociedad civil

El sector periodístico es también un usuario importante en el ecosistema de datos abiertos. Existen medios que conforman sus equipos específicos de periodismo de datos—esto es, unidades que se especializan en el uso de datos como fuente del periodismo, ya sea porque utilizan herramientas digitales para extraer y analizar datos, o bien porque utilizan los datos como herramienta con la que se narra la historia, o una combinación de ambas cosas. Aun cuando algunos medios no cuentan con una unidad especializada, de manera creciente los medios de comunicación utilizan datos abiertos (entre otros datos) para contar una historia convincente por medio del uso de infografías atractivas.

En la región, un caso notable es el equipo de datos de La Nación (La Nación Data) un diario de Buenos Aires, Argentina, que desarrolla una tarea innovadora en el tema de datos abiertos que recibió reconocimiento internacional. En 2013, ganó el Data Journalism Award por haber logrado que se hicieran públicos datos no estructurados, cerrados y opacos de los gastos del Senado argentino, cuya publicación llevó a la iniciación de investigaciones judiciales (puedes ver más información en el siguiente link: <http://blogs.lanacion.com.ar/ddj/data-driven-investigative-journalism/argentina-senate-expenses/>). Otro esfuerzo destacable es el de Poderopedia, una plataforma colaborativa de periodismo de datos en Chile que brinda información acerca de las conexiones entre empresarios y política en el país.

En Uruguay La Diaria ha empezado a trabajar con Periodismo de Datos con su proyecto Río Abierto generando notas periodísticas y un portal de datos producidos y digitalizados por el equipo

La organización de verificación del discurso público Chequeado utiliza en sus investigaciones a los datos abiertos como fuente recurrente. Además han hecho foco particular en comprobar el cumplimiento de los compromisos de apertura de datos realizados por los organismos del Poder Ejecutivo para la aplicación del Decreto 117/16 de Plan de Apertura de Datos.

## 2.3 Empresas, emprendedores y desarrolladores informáticos (hackers), y visualizadores

Este es un grupo muy importante en el reuso y demanda de datos, aunque con diferente enfoque a los anteriores. Como vimos en la unidad uno, existen varias empresas que utilizan datos abiertos como insumo de sus modelos de negocio (por ejemplo, The Climate Corporation). Grandes plataformas también incorporan datos de transporte abierto en la optimización del uso de sus mapas, como es el caso de Google.

Además de las grandes empresas que pueden estar reutilizando datos abiertos, los emprendedores tecnológicos y desarrolladores informáticos (hackers) son quienes suelen estar más presentes en los espacios de innovación cívica y reutilización de datos. Como señalamos en el módulo 2, este tipo de audiencia permite extender las funcionalidades de un portal de datos abiertos ya que poseen conocimientos para usar los datos, procesarlos, y crear nuevos servicios y aplicaciones que permitan crear nuevo valor e innovar con base en las necesidades ciudadanas. Estos emprendedores ven en los datos abiertos un recurso valioso para la innovación social o

comercial que puede tener usos casi ilimitados. Incluimos también en este grupo a los visualizadores—quienes ponen los datos de manera accesible, amigable e incluso bonita para el consumo de los datos. Este grupo es de particular relevancia para fomentar la reutilización de datos en eventos tales como hackathons y desafíos de innovación, ya que su apetito por la innovación sumado a su conocimiento de la tecnología permite la creación de aplicaciones móviles, servicios digitales y plataformas tecnológicas intensivas en datos abiertos (entre otros). Revisaremos más sobre el rol de este grupo en la unidad 4.

## 2.4 Las universidades y la academia

Para el sector académico, los datos abiertos son un insumo para la investigación. Un aporte aún más radical es incluir datos abiertos como recursos integrales en los procesos de enseñanza y que puedan servir de insumos para aportar soluciones innovadoras a problemas sociales. Además de reutilizar los datos la academia también contribuye a la apertura de datos y en algunos casos, incluso a cambios en la normativa. Por ejemplo, en Perú y Argentina, la ley exige que los resultados de las investigaciones financiadas por el Estado estén disponibles en repositorios digitales de acceso público. Esto constituye una nueva dimensión de creación de valor en el ámbito científico a partir de datos abiertos. No solo tiene el potencial de facilitar una mayor participación de la comunidad científica, sino que, además, puede multiplicar los beneficios derivados de la investigación en distintas esferas de la economía, el medio ambiente y la sociedad al hacerlas abiertas y sin restricción a su acceso.

Un ejemplo destacado en la región es la Universidad de Buenos Aires, en Argentina, la cual ha lanzado recientemente su catálogo de datos abiertos, como parte de una política más amplia de “Universidad Abierta”, también destaca a su vez el caso de la Universidad Nacional de Cuyo, que posee un programa denominado “Universidad Transparente” a través del cual promueven las políticas de apertura gubernamental dentro de la Universidad y publican algunos conjuntos de datos y visualizaciones construidas sobre los mismos.

En Uruguay, la Universidad de la República del Uruguay ha realizado varias iniciativas en pro de los datos abiertos y en concreto cuenta con el Núcleo REA con el objetivo contribuir a la inclusión educativa de la población, particularmente de colectivos vulnerables, mediante diversas estrategias y dispositivos basados en recursos digitales, a la vez que se desarrolla investigación, formación de recursos humanos y extensión para la producción y uso de Recursos Educativos Abiertos Accesibles (REAA). Contando con el repositorio Colibrí para la publicación de investigaciones financiadas por la universidad.

La Universidad Católica del Uruguay ha estado experimentando nuevas formas de enseñanza con datos abiertos. El primer experimento de #UcuDataLab se realizó coordinando cursos de Facultad de Ciencias Humanas y Facultad de Ingeniería y Tecnologías con equipos interdisciplinarios para que los alumnos creen artículos de periodismo de datos y prototipos de aplicaciones cívicas incluyendo las visiones de los diferentes cursos. En este momento se está en proceso de formalizar las experiencias

## 2.5 El gobierno

Si bien solemos pensar en el gobierno como el actor principal en la oferta de datos abiertos, el gobierno puede ser otro actor importante en la demanda. Recordemos que uno de los beneficios que los datos abiertos pueden brindar es en mejorar la eficiencia de las operaciones internas, la modernización de los servicios públicos o la mejora en la toma de decisiones de política pública, basada en mejor información.

En este marco, existe una creciente tendencia en la región y el mundo hacia la creación de laboratorios de innovación en gobiernos locales, estatales o nacionales. Estos laboratorios suelen fomentar la demanda al identificar problemas públicos de importancia para organismos de gobierno con soluciones tecnológicas brindadas por su propio equipo, o bien por desarrolladores o hackers cívicos que colaboran con estos laboratorios. Estos laboratorios actúan también como motores para la apertura, ya que la construcción de estas soluciones suelen requerir la colaboración de los propios gobiernos en abrir y aportar datos. Ejemplos en la región incluyen el Laboratorio de Innovación Social en Gobierno Digital de Uruguay, el Laboratorio de Gobierno de Argentina, o el Laboratorio de Gobierno de Chile. En otros casos, el propio gobierno crea aplicaciones web o móviles para el uso de los datos abiertos. Por ejemplo, la Intendencia de Montevideo en Uruguay desarrolló una aplicación de transporte, “Comolr”, para mejorar los traslados en la ciudad. Su versión móvil cuenta ya con más de 100 mil instalaciones.

## 2.6 La ciudadanía (o público en general)

Los ciudadanos constituyen, al menos en teoría, el grupo de usuarios finales de los datos abiertos. Los ciudadanos acceden a los datos abiertos gubernamentales de manera directa o indirecta. En el primer caso, existen quienes acceden al portal de datos para revisar temas de transparencia, como por ejemplo, los salarios de sus representantes. Quizás el canal más común, sin embargo, es el indirecto: es decir, los ciudadanos utilizan o bien consumen datos abiertos gubernamentales mediante el uso de productos o servicios desarrollados por otros actores de la comunidad. Por ejemplo, ciudadanos o ciudadanas interesados en conocer a sus representantes gubernamentales para tomar decisiones en jornadas electorales, podrían consultar el portal Cargografías para entender el pasado político de los funcionarios en su país. Esta iniciativa utiliza datos abiertos y pone en un solo lugar información electoral que de otra manera estaría dispersa en varios sitios e incluso muy difícil de hallar.

Quizá el ejemplo más paradigmático en Uruguay es el de Atuservicio. El Ministerio de Salud Pública (MSP) pone a disposición de los usuarios un conjunto de datos sobre las Instituciones que integran el Seguro Nacional de Salud. Se entiende que dicha información es de interés para los usuarios en general y por eso cada año, la misma ha sido publicada a través de la página web institucional. Sin embargo, esta información no era usada satisfactoriamente por los ciudadanos, por lo que en conjunto con la ONG DATA el MSP genera una plataforma donde pueden visualizarse los datos necesarios para que los ciudadanos tomen una decisión informada acerca de a qué prestador de salud afiliarse.

Otro caso propio de la comunidad de ciudadanos que trabajan con datos públicos en Argentina es el de Data.ar, un portal de datos abiertos no oficial, donde los propios ciudadanos pueden publicar conjuntos de datos de origen gubernamental que no están disponibles en los portales públicos.

No resulta extraño que los usuarios de las apps o servicios construidos con base en datos abiertos no sepan qué son, o de qué se tratan los datos abiertos; para los usuarios, lo importante radica en la funcionalidad y utilidad de estos productos o servicios.

### Introducción

Los datos abiertos abren un mundo de nuevas posibilidades para generar valor, tanto en el ámbito social como en el económico. Como vimos, en el caso de los gobiernos, los datos abiertos pueden ayudar a mejorar la prestación de servicios públicos, facilitar la reducción de costos de operación vía la eliminación de duplicaciones y contribuir a una mejor toma de decisiones de política pública; con mejores datos, y mejor información, las chances de diseñar una mejor política o de prestar un mejor servicios son mayores. El objetivo de esta unidad es presentar dos instrumentos concretos para identificar, promocionar y/o fortalecer a una comunidad de datos abiertos para la mejora de servicios públicos, basados en la colaboración y una buena definición el problema público a resolver.

Ideas originadas en procesos de innovación abierta, como se lo conoce en el sector privado, han permeado a los gobiernos. La mejora de servicios públicos puede ser generada desde adentro del propio gobierno mediante sus procesos de innovación, o bien, mediante espacios de generación de valor público a través de la colaboración y la co-creación. La co-creación para la mejora de servicios públicos se caracteriza por actividades donde al menos dos actores (por ejemplo, público, privado, gubernamental, o cívico) colaboran en la realización de un proyecto para alcanzar determinado resultado.

Como revisamos ya, para generar valor con los datos abiertos, se requiere de la construcción y fomento de comunidades que usen la información y generen aplicaciones de interés para el público. Grupos como los periodistas de datos, visualizadores y organismos de la sociedad civil, pueden ser claves en varias etapas de apertura de datos, ya que a través de ellos, se pueden identificar espacios de mejora en los servicios públicos, crear productos o servicios de interés que generen la atención del público y los propios tomadores de decisiones en el gobierno.

Ahora bien, aun cuando un gobierno—ya sea a nivel local o nacional—cuenta con una política de datos abiertos, puede no resultar fácil identificar en qué mejoras de servicios enfocarse; o de qué manera crear alianzas o colaboraciones dentro y fuera de la organización, para contribuir a la mejora de servicios con datos abiertos.

En esta unidad presentamos dos instrumentos prácticos para avanzar en ambos aspectos:

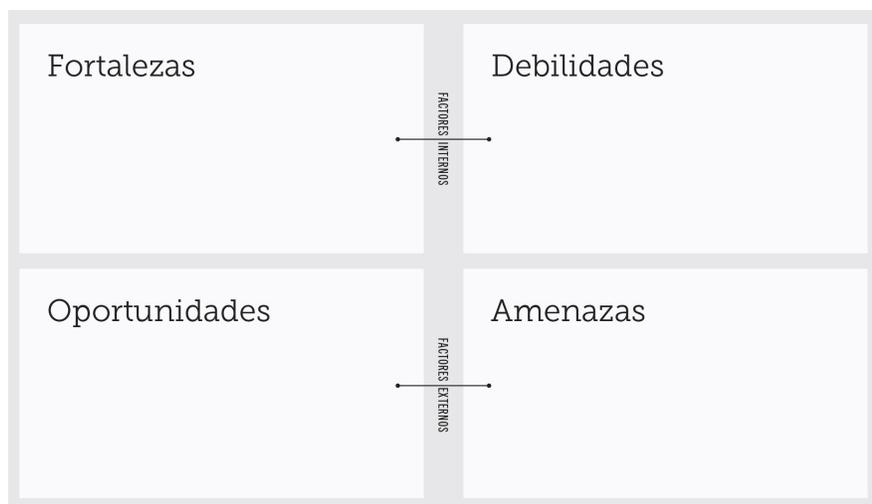
- El instrumento FODA—un análisis de Fortalezas, Oportunidades, Debilidades y Amenazas (las siglas de FODA), para identificar y/o priorizar áreas de mejora de servicios públicos con datos abiertos;
- El mapa de creación de alianzas, para identificar aliados o colaboradores, fuera y dentro del gobierno, para afianzar la colaboración o co-creación de servicios públicos con datos abiertos.

### 3.1 Análisis FODA

FODA es la sigla de Fortalezas, Oportunidades, Debilidades y Amenazas. El análisis FODA es un instrumento que sirve para analizar el estado de situación de una organización, proyecto o sector basado en características internas (fortalezas y debilidades) y la situación externa (oportunidades y amenazas). Para realizar un análisis FODA, entonces, se necesitan identificar y asociar los factores internos y externos que ayudan u obstaculizan lograr un objetivo dado para la mejora de servicios públicos con datos abiertos. El análisis FODA puede ser de utilidad sobre todo en etapas tempranas de apertura de datos y de involucramiento de la demanda, ya que lleva a un entendimiento más amplio de lo que un proyecto y organización pueden ofrecer, cuáles son las debilidades claves sobre las que se debe trabajar para que resulte exitoso, y qué personas externas consultar para ello. Pero es también útil en etapas más avanzadas, para evaluar la viabilidad de un proyecto de mejora para servicios públicos.

La herramienta funciona como autoevaluación, pero incrementa su valor si se usa en diálogo con colegas o en colaboración con otras oficinas relevantes. Se sugiere como herramienta para comenzar un taller interno, por ejemplo, donde se quiera plantear el comienzo de una estrategia de apertura de datos abiertos, o bien, su priorización. Los factores de análisis externo pueden ser de especial relevancia a la hora de identificar a la demanda o comunidad de datos abiertos, y cuáles de sus características actuales pueden determinar la estrategia de datos en cuestión.

A continuación, presentamos la matriz de análisis FODA y sus preguntas guía:



Fuente:  
DIY Toolkit, Nesta:  
<http://es.diytoolkit.org/tools/swot-analysis-2/>

#### **Fortalezas**

¿Qué haces para hacer las cosas mejor que los demás?  
¿Qué te hace único?  
¿Qué recursos únicos o de bajo costo puedes usar que los demás no?  
¿Cuáles son las cosas que las personas en tu sector ven como tus fortalezas?

#### **Oportunidades**

¿Qué necesitan las personas?  
¿Prefieren otra cosa?  
¿Hay algún avance en la tecnología?  
¿Hay cambios en las políticas del gobierno?

#### **Debilidades**

¿Qué podrías mejorar?  
¿Qué debes evitar?  
¿Cuáles son las cosas que los usuarios podrían ver como debilidades?

#### **Amenazas**

¿Qué amenazas enfrentas?  
¿Qué están haciendo tus competidores?  
¿Es particularmente difícil un escenario con tecnología cambiante?  
¿Hay restricciones financieras importantes?

### 3.2 Mapa de creación de alianzas

El mapa de creación de alianzas es de especial relevancia para trabajar de manera colaborativa y por ende, para la identificación o inclusión de actores en una estrategia de mejora de servicios con datos abiertos. Incluir usuarios de datos abiertos en las etapas iniciales de una estrategia de mejora es una buena forma no solo de incrementar la capacidad de trabajar con estos datos, sino también de aumentar su alcance y potencial impacto. Las alianzas con actores de la comunidad ayudan a construir un entendimiento común y fortalecer y expandir el conocimiento del tema a través de varias y diferentes perspectivas.

Generar estas colaboraciones o alianzas puede resultar en una inversión considerable de tiempo, más aún, cuando en el sector público, las normas y regulaciones existentes, o la propia cultura organizacional pueden ser obstáculos no menores para poner actores externos en la mesa de diálogo.

El Mapa de Creación de Alianzas puede ser útil, ya que separa el proceso de colaboración en 12 pasos, lo cual contribuye a anticipar dificultades y quitarle complejidad. El mapa es una herramienta auto-administrada bastante simple que, al igual que en el análisis FODA, puede ser utilizada en un taller con colegas para maximizar su utilidad. El Mapa de Creación de Alianzas puede ser usado para analizar en qué fase de alianza se encuentra la organización. Por ejemplo, en una primera etapa, los 5 primeros pasos resultan claves. El mapa puede servir para revisiones conjuntas en etapas más avanzadas, una vez que se han creado colaboraciones con socios.



Fuente:  
DIY Toolkit, Nesta:  
<http://es.diytoolkit.org/tools/building-partnerships-map-2/>

#### Doce pasos para un Mapa de Alianzas:

1. **Determinar:** implica entender el desafío, recolectar información, consultar con las partes interesadas y con proveedores potenciales de recursos externos, creando una visión de/para la alianza.
2. **Identificar:** implica identificar socios potenciales y, de ser apropiado, asegurar su participación, motivándolos y alentándolos a trabajar juntos.
3. **Construir:** Los socios construyen su relación de trabajo estando de acuerdo con sus metas, objetivos y principios fundamentales que forman la base de su alianza.
4. **Planear:** Los socios planean un programa de actividades y comienzan a bocetar un proyecto coherente.
5. **Dirigir:** Los socios exploran la estructura y dirección de su alianza, de mediano a largo plazo.
6. **Asignar recursos:** Los socios (y otros partidarios) identifican y movilizan dinero y/o recursos no monetarios.

7. Implementar: Una vez que los recursos están en su lugar y se han acordado los detalles del proyecto, empieza el proceso de implementación; trabajar sobre un horario pre-acordado y bajo metas específicas.
8. Medir: Medir y reportar sobre la efectividad y el impacto, producción y resultados. ¿Está cumpliendo sus metas la alianza?
9. Analizar: Analizar la alianza, ¿cuál es el impacto de la alianza en organizaciones asociadas? ¿Es tiempo de abandonar algunos socios/o que nuevos socios se unan?
10. Revisar: Revisar la alianza, programa(s) o proyecto(s) bajo la luz de la experiencia.
11. Institucionalizar: Construir una estructura apropiada y mecanismos que aseguren el compromiso y la continuidad de la alianza.
12. Mantener o interrumpir: Crear sustentabilidad o acordar una conclusión apropiada.

Los 12 pasos que presenta el mapa pueden resultar algo abstractos. Para un ejemplo de cómo se utilizó el mapa de alianzas para un proyecto de mejora de salud en la India, revisa este link: <http://es.diytoolkit.org/casestudies/595>

# Datos abiertos para la economía: innovación, emprendedores y datos

## 4.1 Introducción

A pesar de las expectativas, alcanzar los beneficios esperados de los datos abiertos es complejo. Parte del problema que enfrentan los gobiernos es la necesidad de cultivar un ecosistema de reutilizadores. Por el lado de la demanda, la autoridad responsable de la plataforma de datos abiertos necesita cultivar una base de reutilizadores independientes, ya sea para que contribuyan con nuevos servicios o aplicaciones basados en datos abiertos, así como de usuarios finales que consumen estos servicios. Un problema clave que enfrentan las autoridades de las plataformas de datos abiertos es de gobernanza: es decir, cómo atraer y mantener activos a quienes tienen la capacidad de generar nuevos servicios y aplicaciones para la economía. Esta unidad introduce un modelo simple para fomentar la innovación con datos abiertos, basado en reglas y herramientas.

La lógica de la contribución de los datos abiertos a la generación de beneficios económicos se resume en dos tipos:

**Beneficios directos:** incluyen la generación de ingresos a partir de nuevos productos o servicios, la creación de puestos de trabajo o la reducción de costos económicos. Un ejemplo sería una nueva compañía que obtiene sus ingresos a partir de visualizar o transformar datos abiertos para la industria financiera.

**Beneficios indirectos:** la reducción de tiempos de espera para usuarios de aplicaciones con datos abiertos, el incremento en opciones de búsqueda para la toma de decisiones, el aumento en la eficiencia del sector público y el crecimiento de industrias relacionadas, entre otros. Un ejemplo de beneficios indirectos sería la reducción de horas perdidas en el tráfico por los habitantes de una región, a partir del uso de datos abiertos en el transporte.

Para cultivar un ecosistema de emprendedores e innovadores alrededor de los datos abiertos, el enfoque de plataforma de innovación es de utilidad, el cual desarrollamos a continuación.

## 4.2 Plataformas de innovación digital y el caso de datos abiertos

Las plataformas de innovación digital actúan como fundamentos sobre los cuales otras empresas pueden construir productos, servicios o tecnologías complementarias. Las plataformas de innovación digital se caracterizan por tener una “autoridad responsable”, que es quien controla la innovación en la arquitectura central, así como las actividades de innovación que suceden en el ecosistema de innovadores externos. Un ejemplo de plataforma de innovación es el sistema operativo móvil Android, cuya funcionalidad se basa en un ecosistema de desarrolladores que crean nuevos servicios mediante la innovación de aplicaciones móviles.

Las plataformas de datos abiertos pueden ser analizadas como plataformas para la innovación, ya que ponen a disposición conjuntos de datos a terceros, a partir de los cuales pueden generarse nuevos servicios o productos.

Al igual que en el caso de las plataformas de innovación digital del mundo comercial, cultivar el crecimiento de un ecosistema de innovadores y desarrolladores es un papel crucial de la autoridad de la plataforma de datos abiertos. La capacidad de cultivar innovadores dentro de un ecosistema de plataforma es esencialmente una tarea que consiste en habilitar aquellas actividades de los desarrolladores que agregan valor a la plataforma, y a la vez restringir aquellas que quitan valor.

#### **4.3 Reglas y herramientas para cultivar la innovación desde la demanda**

En términos simples, la teoría de innovación en plataformas digitales del mundo comercial, sugiere que para cultivar un ecosistema de innovadores se requieren dos elementos:

**Reglas:** aquellas normas proporcionados por la autoridad de la plataforma que autoriza o restringe a los miembros del ecosistema a realizar determinadas acciones, con el fin de asegurar su integridad. Por ejemplo, en los casos comerciales como el sistema operativo Android o Apple iOS, las reglas toman la forma de licencia o términos y condiciones a los cuales los desarrolladores deben sostenerse, tales como la prohibición de desarrollar apps con fines criminales.

**Herramientas:** son aquellos artefactos proporcionados por la autoridad de la plataforma que habilitan recursos y permiten a los desarrolladores generar nuevos productos o servicios. En el mundo comercial, ejemplos de herramientas son las interfaces de programación de aplicaciones (API), que proporcionan a los desarrolladores acceso a la a la plataforma central, y los kits de desarrollo de software (SDK), que facilitan a los desarrolladores la creación de servicios de software tales como apps.

#### **4.4 Reglas para fomentar la innovación con datos abiertos**

Tal como fuera anticipado, las reglas son importantes a la hora de fomentar la innovación por parte de terceros, porque permiten mantener la integridad de la plataforma. En el caso de las plataformas de datos abiertos, las reglas toman básicamente dos formas:

**Términos y condiciones de uso:** suele ser una regla clave para mantener la integridad de la plataforma; en general, los términos y condiciones explicitan responsabilidades, y a la vez expresan aquellos usos permitidos, tales como fines lícitos.

**Licencias:** generalmente, los datos abiertos incluyen licencias que permiten la reutilización. Por ejemplo, el Gobierno de la Ciudad de Buenos Aires aplica la Licencia Creative Commons Atribución 2.5 Argentina (CC BY 2.5 AR) para todos sus contenidos digitales y recomienda su uso como una de las reglas para innovación por parte de terceros.

#### **4.5 Reglas para fomentar la innovación con datos abiertos**

En todos los casos, para incrementar la utilidad de estas herramientas, es importante trabajar con una serie de prioridades o problemas públicos a resolver con datos abiertos.

A fin de cultivar la demanda de emprendedores e innovadores con datos abiertos, entre las varias herramientas sociales que pueden utilizarse, se puede distinguir entre dos tipos de herramientas: i) información y software, y ii) sociales.

### **Herramientas de información y software**

Estos son los recursos más obvios en una plataforma de datos abiertos, y comprenden a los conjuntos de datos, así como otras herramientas de información, como manuales y videos relacionados para su uso. También incluye herramientas de software como APIs y portales web que permiten a los desarrolladores e innovadores acceder a conjuntos de datos de manera programática.

### **Herramientas sociales**

Las herramientas sociales complementan a las de información y software, y son aquellas que involucran la participación de la demanda, mediante el despliegue de eventos, financiación u otro tipo de incentivos. Se destacan tres:

**Hackatones:** estas maratones de ideas, diseño y prototipado de software, reúnen a especialistas, técnicos programadores, diseñadores y emprendedores para trabajar de manera colaborativa. Las jornadas suelen durar entre 24 y 48 horas y suelen tener desafíos específicos para guiar el evento, y en muchos casos cuentan también con talleres paralelos para la creación de capacidades. Las hackatones suelen ser de las primeras estrategias para fomentar la re-utilización de datos abiertos—y a la vez, crear o fortalecer la comunidad, sobre todo, de emprendedores. Por ejemplo, la Ciudad de Buenos Aires utilizó hackatones como sus primeras estrategias para la reutilización de datos. La primera hackaton tuvo lugar en el 2012, y fue la primera organizada por una instancia gubernamental en Argentina. En la Ciudad de México, la realización del primer HackDF—hackaton a principios del 2014—también funcionó como un primer canal para acercarse a la ciudadanía, y a la vez fortalecer el equipo del LABplc (varios de sus ganadores fueron después reclutados para trabajar en el LAB).

**Competencias de aplicaciones (Apps Challenges):** Las competencias son similares a los hackatones, pero habitualmente destacan el elemento competitivo con premios o reconocimiento otorgados por un panel de expertos o jueces, y pueden tener una duración mayor a una hackaton. Existen reglas de la competencia, donde se especifican los productos o aplicaciones considerados “admisibles” para los premios. Ejemplos incluyen las competencias de aplicaciones de la ciudad de Buenos Aires, y la iniciativa regional Desarrollando América Latina que funcionó desde el 2012 al 2016, la cual paso de ser una hackaton regional a una apps challenge con actividades de “mentoring” y seguimiento a proyectos (mas información: <https://es.globalvoices.org/2013/09/26/desarrollando-america-latina-2013-un-apps-challenge-para-aplicaciones-de-impacto-social/>).

**Procesos de incubación y fondos semilla:** estas herramientas se utilizan para incubar proyectos bien formulados que tengan potencial para el desarrollo de una aplicación, producto o servicio basado en datos abiertos, y para ello, se ofrece financiamiento semilla para construir y aplicar modelos ampliables. Un ejemplo de un programa de competencia con incubación fue el concurso de internet de las cosas desarrollado por el Laboratorio de innovación de la ciudad de Buenos

Aires (BA IoT). El BA IoT intentaba afianzar una red de innovadores—internos y externos—que puedan usar tecnología, datos y aplicaciones digitales para mejorar la calidad de vida de la ciudad, mediante una serie de talleres, incubación y premiación, a lo largo de 7 meses. Aun cuando estos programas pueden tener mayor éxito en cuanto a generar impacto con datos abiertos, la realidad es que son muy pocos los casos en América Latina que cuentan con estos programas.

### **Hackatones: seis aspectos claves a tener en cuenta para organizarlos**

Es importante destacar que las hackatones son poco propensos a generar sostenibilidad o continuidad de los proyectos en el tiempo. Es por ello que varios gobiernos utilizan las hackatones en conjunto con otras estrategias, tales como los Apps Challenges o programas de incubación de iniciativas. Un hackaton puede ser de utilidad como una herramienta para fomentar la innovación especialmente en etapas tempranas de estrategias de datos abiertos. Cabe destacar que los hackatones han sido sobre-utilizados, y en muchos casos realizados sin un buen plan, una buena agenda o alianzas estratégicas, lo cual ha mermado su valor.

Un hackaton no es la panacea de la re-utilización de datos. Por ello, vale la pena destacar brevemente seis aspectos a considerar en caso de decidir organizar uno:

1. Definir un problema específico sobre el cual trabajar: plantear un buen reto que conecte un problema público con las capacidades de los hackers y participantes para trabajar en su solución no es sencillo. Una regla para reducir su complejidad es elegir un tema, presentar los problemas de forma desagregada, o incluso trabajar con una agencia en particular para la solución de un problema específico. Por ejemplo, el hackaton verde de la Ciudad de Buenos Aires (2014, ver multimedia introductorio), se enfocó en buscar soluciones a problemas de medio ambiente.
2. Trabajar en las fuentes de datos e información: parte de definir los retos de un hackaton está determinado por la información disponible, últimos avances y retos en cada área que puedan poner en contexto a los participantes sobre lo que se busca resolver en el evento. Los datos abiertos que se utilicen en el evento deben ser de calidad y relevancia para que los participantes puedan hacer diagnósticos más precisos de los problemas a los que se enfrentan y puedan llenar los prototipos de aplicaciones con información lista para consumir.
3. Trabajar con aliados: esta es la oportunidad para conectarse e invitar a actores de la comunidad que puedan ser colaboradores estratégicos en resolver un problema. Cada actor ofrece algo que aportar, y trabajar con expertos, organizaciones e individuos cercanos al tema de interés resulta clave; asimismo, contar con el apoyo de proveedores de datos, especialistas en tecnología y medios de comunicación puede marcar la diferencia en los resultados del evento. El mapa de alianzas que presentamos en la unidad 3 puede ser de utilidad para este fin.
4. Planear la realización del evento a detalle: los preparativos de un hackaton incluyen la identificación de un equipo a cargo, la selección de un día, lugar y horarios, trabajar con aliados y promocionar el evento. La “Open Data Hackathon How to Guide” (solo disponible en inglés) cubre en detalle estos aspectos.

5. Diseñar la agenda del evento: un hackaton no es un evento totalmente desestructurado; si bien, gran parte del tiempo se dedica al trabajo en grupos, una agenda clara que dé forma al proceso creativo es de gran importancia. Conocer los problemas, crear un equipo, decidir una idea, planificar el tiempo y preparar una presentación son tareas necesarias en un hackaton que se pueden facilitar con una buena agenda.
6. Utilizar mecanismos de seguimiento y tomar nota de lo aprendido: Dar seguimiento a un hackaton puede resultar clave para que los esfuerzos no se olviden o mueran una vez que termina el evento. Idealmente, un hackaton se puede combinar con un programa de incubación o seguimiento que contribuya a amplificar los resultados. En otros casos, un hackaton puede ser una oportunidad para hallar talento que pueda ser luego incorporado en el propio gobierno. Por otra parte, tomar nota de lo aprendido es sumamente relevante. Tomar nota de todo lo que resultó bien, para repetirlo en un próximo evento, así como de lo que salió mal, para evitarlo. Realizar una breve encuesta una vez terminado el evento a los participantes es también una buena idea para aprender lecciones.

# Datos abiertos y políticas públicas: Desafíos y aspectos críticos de los datos abiertos y su demanda

## 5.1 Introducción

Generar espacios para la co-creación, conocer a aquellos que demandan datos abiertos y los roles de los actores en el ecosistema, resulta clave a la hora de potenciar el valor de los datos abiertos en un contexto determinado. En esta unidad, el objetivo es reflexionar sobre lo aprendido en el módulo, y presentar una serie de desafíos y elementos a tener en cuenta en el diseño de políticas públicas basadas en datos abiertos gubernamentales, con énfasis en la demanda de datos.

## 5.2 Desafíos, futuro y reflexión

Enfocarse en la calidad de los datos abiertos. Como se revisó en módulos anteriores, es importante priorizar la calidad de los datos—y no solo la cantidad de los conjuntos de datos que se ponen disponibles en formato abierto. Por ello, proponer lineamientos lo suficientemente flexibles pero a la vez comprometidos con la calidad de los datos es clave. La calidad no asegura el uso de los datos; pero aumenta el potencial de creación de valor público en ámbitos locales, estatales y nacionales.

Incluir la perspectiva de género en el diseño de políticas de datos abiertos. Un aspecto crítico para contribuir a la generación de valor con datos abiertos, implica mantener un foco de inclusión. A pesar de iniciativas a nivel global, como la de Gobierno Abierto Feminista, las mujeres y otras minorías siguen siendo excluidas de las políticas de datos, no solo a la hora de publicar datos (brecha de género), sino también al incluirlas como parte de la demanda. Entre los compromisos de la Alianza de Gobierno, por ejemplo, tan solo un 2% de ellos incorporan una perspectiva de género. En datos abiertos, el hecho de que no haya datos desagregados por género significa que hay un mundo que desconocemos, con el potencial para contribuir a mayores inequidades. Por ejemplo, es sabido en la región que existe una carencia general de datos sobre femicidios que afectan tanto a las mujeres como a otras comunidades marginadas. Para ello, es clave incorporar a sectores de la demanda en el diseño de políticas con datos abiertos, tales como asociaciones por los derechos de la mujer, de derechos humanos o personas vulnerables.

Difundir y documentar. la medición de impacto de los datos abiertos es aún incipiente. El uso de videos cortos y blogs desde el gobierno suelen ser buenas ideas para documentar usos y promover la difusión de casos incipientes con datos abiertos. Se debe invertir de inmediato en conocer mejor quién usa los datos, para qué y con qué resultado. En muchos casos, las propias organizaciones que abren datos desconocen las métricas básicas de uso o demanda de datos abiertos.

### 5.3 Fomentar la demanda

Desarrollar casos de uso con datos abiertos. Para fomentar la demanda, se requiere mostrar los beneficios que generan los datos abiertos. Como observamos en módulos anteriores, los aspectos tanto técnicos como políticos son de gran relevancia para establecer políticas basadas en datos abiertos. Para ello, la recopilación de casos de uso detallado sobre datos abiertos resulta clave. Los casos deben permitir explorar a mayor profundidad los aspectos socio-técnicos del sistema—es decir, cómo se desarrolla y evoluciona un proyecto de datos abiertos, cómo se manejan sus intereses, como se desenvuelven las complejas relaciones entre actores, intereses y aspectos contingentes en acción. Para ello, concebir alianzas con académicos, periodistas o equipos de evaluación de políticas puede ser útil para sistematizar casos de uso.

Fomentar más canales de colaboración entre ciudadanía y gobierno. Abrir datos sin una comunidad que los use no generará valor per se. Es común encontrar que funcionarios públicos señalan su dificultad para priorizar aquellos datos que puedan ser de mayor interés a los grupos que componen la demanda de datos. Facilitar espacios de colaboración entre gobierno y esos actores es una estrategia que ha probado ser efectiva a la hora de crear y fomentar el uso de datos abiertos. Tal como hemos señalado en la unidad 4, los hackatones o concursos de apps han demostrado utilidad, pero sería deseable invertir en proyectos de incubación, o de seguimiento a más largo plazo.

Invertir en entender los modelos de negocios con datos abiertos. Si bien los emprendedores tecnológicos, así como las startups y empresas del sector privado son actores muy relevantes en el ecosistema de datos abiertos, siguen estando desconectados en la región, o incluso ausentes. Hacer más atractivo el involucramiento de emprendedores es un desafío que debería ser abarcado con mayor énfasis. No es fácil generar negocios con recursos abiertos—por eso, se necesita también hacer hincapié en entender mejor qué tipo de modelos de negocios son viables o sustentables en el ecosistema.

Los esfuerzos por estimular la co-creación de valor con datos abiertos no deben ser restringidos a políticas del tipo “arriba-hacia-abajo” (top-down). Si bien el liderazgo y el apoyo político resultan claves en el proceso de abrir datos de gobierno, los esfuerzos que nacen desde organizaciones de la sociedad civil, o de los mandos medios del gobierno, son igualmente importantes—tal como lo demuestran varios de los ejemplos de este módulo.

Crear más espacios de colaboración internos: las unidades de innovación de datos abiertos, muestran que son catalizadoras, no creadoras o proveedoras de servicios públicos. La creación de capacidades internas así como la generación de cambio cultural en la apertura de información siguen siendo aspectos necesarios, y que requieren de la inversión de recursos. Alianzas con programas como el fellowship de la Escuela de Datos, o de programadores cívicos de la iniciativa de Code for America, pueden ayudar a potenciar el uso de datos por medio del trabajo conjunto con usuarios internos.

## 5.4 Ética y redes de cooperación

Cuidar la ética de los datos abiertos. A pesar de su gran potencial para hacer el bien, los datos abiertos también pueden ser utilizados para hacer el mal, como sucede con todas las tecnologías. En los esfuerzos por la apertura de datos, se deben tener especialmente en cuenta los riesgos de quienes pueden ser más vulnerables a una discriminación o mal uso de datos. Divulgar las identidades de los ciudadanos o sus datos contextuales sin su consentimiento, por ejemplo, puede generar daños, discriminación o un trato injusto. Un buen ejemplo de este problema fue la polémica app llamada "rastreador de ghettos" lanzada en septiembre de 2013 en los Estados Unidos. Esta app se creó, supuestamente, para ayudar a los usuarios a orientarse por las distintas zonas de la ciudad y encontrar las zonas seguras, pero generó críticas inmediatamente debido a sus estereotipos racistas y clásicos de lo que es bueno y malo. Los algoritmos no se crean en un vacío técnico, y los datos abiertos son cualquier cosa menos objetivos. Es crítico encontrar el equilibrio entre las iniciativas de protección de datos y los beneficios de la transparencia y rendición de cuentas derivados de los datos abiertos. Los formuladores de políticas y los legisladores deben encontrar la manera de obtener beneficios de los datos abiertos sin afectar las necesidades de privacidad de los individuos.

Facilitar y fomentar las redes interregionales de innovadores con datos abiertos. Sin dudas, ciudades en la región comparten dinámicas y necesidades similares. Compartir experiencias y aprender de quienes ya han avanzado representa un valioso recurso que se debe fomentar. Crear una red de innovadores en datos abiertos con otras regiones del país o incluso otros países de la región, puede contribuir a reforzar aprendizajes y minimizar errores o riesgos en el corto plazo.

## MÓDULO 3

# Lecturas

Aguerre C., Bonina, C. (2019). Por Mi Barrio. Lecciones, Hallazgos y Futuro, Documento de Trabajo. Recuperado de: <http://doi.org/10.5281/zenodo.2531028>

Bonina, C. (2015). Co-creación, innovación y datos abiertos en ciudades de América Latina: lecciones de Buenos Aires, Ciudad de México y Montevideo, Documento de Trabajo ILDA. Recuperado de <http://riga.idatosabiertos.org//papers/co-creacion-innovacion-y-datos-abiertos-en-ciudades-de-america-latina-lecciones-de-buenos-aires-ciudad-de-mexico-y-montevideo.html>

Bonina, C. (2018). El valor del open data en América Latina. En Economía de los Datos Riqueza 4.0. Madrid: Fundación Telefónica. Recuperado de <http://doi.org/10.5281/zenodo.1310030>

Patiño, J. (Ed.) (2014). Datos Abiertos Y Ciudades Inteligentes En América Latina: Estudios de Casos, Comisión Económica para América Latina y el Caribe. Recuperado de: <https://www.cepal.org/es/publicaciones/37089-datos-abiertos-ciudades-inteligentes-america-latina-estudio-casos>

Innovando para una mejor gestión: La contribución de los laboratorios de innovación pública, Acevedo, Sebastián; Dassen, Nicolás: <https://publications.iadb.org/es/publicacion/17174/innovando-para-una-mejor-gestion-la-contribucion-de-los-laboratorios-de>

Prueba estas 5 metodologías de innovación abierta - Blog Abierto al Público: <https://blogs.iadb.org/conocimiento-abierto/es/5-metodologias-de-innovacion-abierta/>

¿Qué sucede después de un Hackatón? - Blog Abierto al Público: <https://blogs.iadb.org/conocimiento-abierto/es/que-sucede-despues-de-un-hackaton/>

Sección Código para el Desarrollo - Blog Abierto al Público: <https://blogs.iadb.org/conocimiento-abierto/es/codigo-para-el-desarrollo/>

Sección de Código Abierto - Blog Abierto al Público: <https://blogs.iadb.org/conocimiento-abierto/es/codigo-abierto/>

Caso: BA Obras - Blog Abierto al Público: <https://blogs.iadb.org/conocimiento-abierto/es/ba-obras-plataforma-obras-publicas-abiertas/>

Caso: San Pedro Obras - Blog Mejorando Vidas <https://www.iadb.org/en/improvinglives/open-source-code-crosses-borders> <https://www.iadb.org/es/mejorandovidas/codigo-abierto-que-traspasa-fronteras>

Caso: Consul Madrid - Blog Abierto al Público: <https://blogs.iadb.org/conocimiento-abierto/es/consul-ciudad-abierta-codigo-abierto-aeriales.htm>

## MÓDULO 4

---

# Desarrollando habilidades prácticas para trabajar con datos

---

Docente: Juan Casanueva  
Contribuciones: Javiera Atenas  
Camila Salazar  
Daniel Villatoro  
Sergio Araiza

TUTORIAS



Manuel Podetti  
Silvia Da Rosa



Romina Colman  
Susana Soto

# Resumen Introdutorio

Este módulo se centra en desarrollar competencias que les permitan usar herramientas para enfrentar de manera práctica los retos que implica el trabajar con datos de manera cotidiana, en específico, para poder limpiar, analizar y visualizar datos abiertos. Las secciones del módulo son las siguientes:

- Trabajando con herramientas de visualización: Introducción al desarrollo de gráficos y de infografías para representar visualmente información numérica en reportes o informes.
- Trabajando con Georreferenciación: Introducción a los conceptos básicos de la georreferenciación con tal de poder manejar datos referenciados en las coordenadas en un mapa.
- Calidad de los datos: Introducción a la técnicas de limpieza de datos para asegurar la calidad de datos, para evitar que los datos que se publiquen sean poco claros para quien los quiera utilizar, evitando la presencia de datos incorrectos o inexactos.
- Una mirada básica a estadísticas y datos: Introducción a conceptos básicos de estadísticas (promedio, media, moda, mediana, correlaciones) para el análisis de datos abiertos.
- Introducción al periodismo de datos: Introducción a las técnicas del manejo de datos y narrativas basadas en datos dentro de diversas corrientes del periodismo de datos

## OBJETIVOS DE APRENDIZAJE

1. Identificar las características que hacen a una base de datos “limpia”, es decir, utilizable sin presencia de errores en la estructura y formatos de los datos
2. Conocer los principales métodos y herramientas para realizar limpieza de datos
3. Conocer (y saber diferenciar) conceptos básicos de estadística
4. Aplicar de manera práctica funciones y herramientas básicas de hoja de cálculo para el análisis de datos
5. Identificar los elementos básicos de los datos geo-referenciados
6. Aplicar de manera introductoria habilidades para el uso y manejo de mapas
7. Identificar las características de distintos enfoques del periodismo de datos

## GLOSARIO DE TÉRMINOS

**API:** El acrónimo de application programming interface se refiere a un conjunto de herramientas de programación que facilitan la comunicación entre softwares. A través de un API se puede acceder directamente desde un programa de análisis de datos, app o página web los datos de una plataforma de datos abiertos.

**Conjunto de datos:** Un conjunto de datos es el agrupamiento de recursos de datos sobre una misma temática.

**Dato:** Un dato es una representación formalizada de entidades o hechos, presentada de forma adecuada para la comunicación, interpretación y procesamiento por medios humanos o automáticos.

**Dataset:** Un dataset es una colección de bases de datos, usualmente del mismo tema, que por su tamaño o característica se entrega en varios archivos de base de datos para facilitar su procesamiento

**Formato CSV:** Los archivos en formato CSV o en inglés comma-separated values son archivos en formatos abiertos de texto plano que almacena datos en forma de tabla, en las que las columnas se separan por comas y las filas por saltos de línea.

**Formato JSON** – JavaScript Object Notation es un lenguaje muy sencillo de automatizar su procesamiento. Si bien no está tan expandido como XML su filosofía similar y su fácil interpretación lo hacen un lenguaje muy apropiado para el intercambio de información.

**Formato KML** – Los archivos bajo el formato Keyhole Markup Language son utilizados para la representación de datos geográficos en tres dimensiones, basado en XML. Este formato se utiliza para guardar datos vectoriales pero también puede contener en archivos con terminación .kmz imágenes sobrepuestas, íconos y otros elementos.

**Formato XLS** – Es el formato de archivo creado por Microsoft para su herramienta de hoja de cálculo.

**Formato XML** – Es el formato extensible de marcas, lenguaje que permite la definición de etiquetas o marcas derivado del SGML.

**Formato Shapes:** Es el formato más popular de los archivos vectoriales. Este es un archivo que fue desarrollado originalmente por la empresa ESRI y que puede contener únicamente un tipo de dato vectorial (o puntos, o líneas o polígonos). Los archivos shape realmente son un conjunto de archivos binarios que para desplegarlos deben tener el mismo nombre y ubicarse en un mismo folder. Los tres archivos básicos que componen un shapefile terminan en .shp, .shx y .dbf.

**Formato GeoJSON:** Este formato de datos geoespaciales es uno de los formatos abiertos más utilizados para distribuir datos geográficos es el GeoJSON. Este, al igual que el JSON popularmente utilizado en el web, es un archivo de texto pero que también guarda las coordenadas de los vértices de las entidades que se representan. Al ser un archivo de texto, en geometrías complejas (con muchos vértices) su tamaño será mucho mayor que cuando los datos se guardan en formatos binarios .

**Formato GeoPackage:** Este es un formato abierto que ha tomado fuerza en la comunidad de usuarios de SIG es el GeoPackage (terminación de archivos en .gpkg). Este es un archivo binario (es un contenedor de la base de datos SQLite) que tiene la gran ventaja de poder contener tanto datos vectoriales como raster e inclusive tablas de metadatos.

**Formato GeoTIFF:** En el caso de los datos raster, existen una gran variedad de archivos capaces de almacenarlos. Uno de los más populares es el GeoTIFF, una modificación al popular archivo de imágenes TIFF, pero que le permite guardar datos con coordenadas geográficas. Normalmente el archivo con terminación .tif debe de acompañarse con un archivo del mismo nombre pero con terminación .tfw, que es el world file que le da la referencia geográfica al raster.

**Georreferenciación:** Es la conversión de localizaciones que los humanos pueden entender en números y códigos que las computadoras pueden entender, como longitud y latitud.

**Metadatos (metadata):** Los Metadatos son datos o descripciones estructuradas que describen las características de otros datos o informaciones.

**PDF** – Es el formato portable de documentos, estándar para la publicación de documentos en la web.

**TXT** – El el formato de archivo de texto plano.

**Infografía:** Las infografías son piezas de material gráfico (digital o impreso) que tiene como objetivo fundamental,comunicar información, que en primer orden parece compleja, de formas simples a usuarios con intereses definidos.

**Webservice**– En el caso que la información sea extraída de una fuente de información viva, en constante modificación; se recomienda la utilización de servicios web que expongan dicha información en tiempo real.

### Introducción

La visualización de datos mantiene una profunda conexión con la necesidad de compartir y comunicar el valor que los datos tienen cuando son aplicados a diversas circunstancias. Para esto, la visualización de datos se apoya en diversas técnicas y métodos para simplificar temas cargados de tecnicismos y transformarlos en contenidos más amigables y dosificados de acuerdo al grado de especialización temática de la audiencia.

Comunicar datos a través de visualizaciones atractivas representa uno de los esfuerzos más amplios del proceso de datos ya que conlleva el conocimiento específico de un área temática más la creatividad necesaria para conectar con diversos tipos de audiencias y sus necesidades específicas.

### Visualización de datos

Trabajar con datos no se limita únicamente a la capacidad de interpretarlos, también implica un fuerte componente de comunicación, ya que los resultados encontrados en la etapa de análisis deberían, idealmente, ser comunicados a más personas o grupos de trabajo de tal forma que se genere una red de conocimiento basada en información proveniente de fuentes en datos abiertos.

A la hora de comunicar los resultados del análisis otro reto se avista en el horizonte; el reto de convertir los resultados (números) en objetos fácilmente interpretables para diversas audiencias. Estas audiencias pueden contar con un legítimo interés por la información pero posiblemente tengan poco tiempo o deseo para realizar un análisis de datos propio. Es por razones como esta que la comunicación visual de datos representa hoy en día uno de las etapas más importantes en el manejo de datos, ya que permite transmitir el cúmulo de conocimiento adquirido a diversos grupos sin la complejidad que representa todo el proceso de manejo de datos.

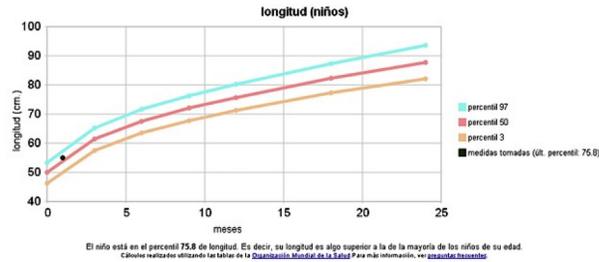
La comunicación visual de datos no debe interpretarse como un esfuerzo nuevo o reciente, en realidad lleva años perfeccionándose en diversas áreas de aplicación como lo son los deportes, las finanzas y el sector económico, la física o el periodismo. Todas estas disciplinas trabajan con diversos enfoques para resolver una pregunta en común: ¿cómo comunicar mejor los datos?

## Tipos de gráficos

Un tipo de gráfico para cada tipo de dato: Para elegir cómo visualizar los datos es necesario valorar bien su naturaleza y el objetivo que se quiere alcanzar con la visualización.

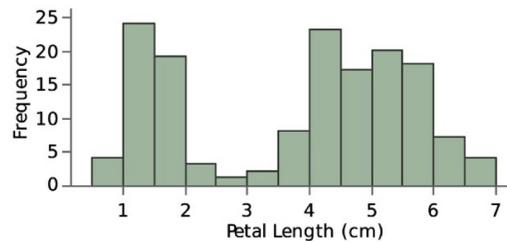
### Gráfico de líneas

Los gráficos de líneas muestran los valores en un intervalo permitiendo ver cómo se ha desarrollado durante el período de la medición.



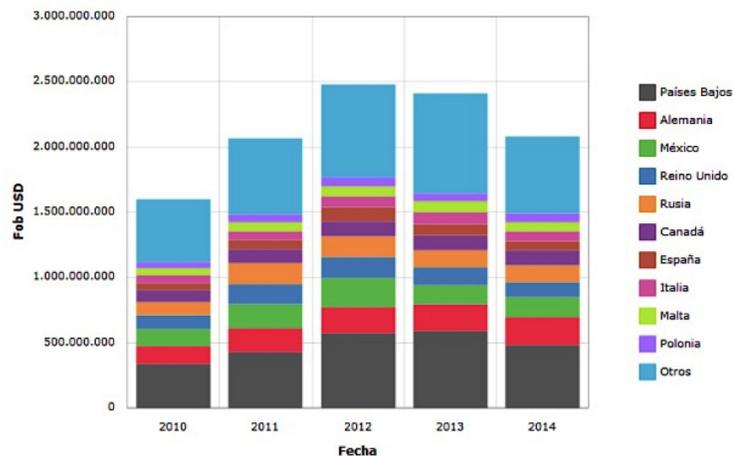
### Histogramas

Cuando los valores de cada elemento, como en las naciones del ejemplo, no presentan una relación recíproca unos con otros, o cuando es importante evidenciar algunos valores específicos, es preferible un histograma a un gráfico de puntos o líneas.



### Gráfico de barras

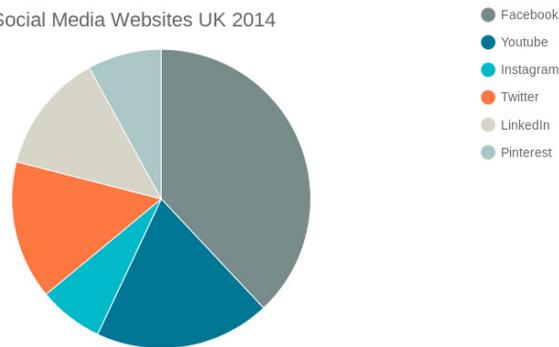
Este tipo es muy útil si se tienen muchas etiquetas largas para mostrar. Aquí la información se muestra en la parte izquierda de la visualización, de manera que conseguimos mayor espacio para que la lectura de los datos sea más cómoda y fácil.



## Gráfico circular (Pie chart)

Se utiliza solo para mostrar datos que sumados entre sí representan un total (y tomados singularmente representan las partes de un total).

Most Popular Social Media Websites UK 2014

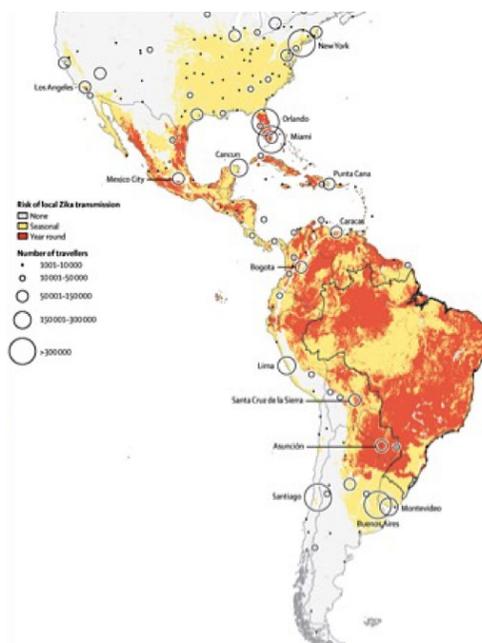


## Mapeando Datos

Visualizar los datos en un mapa requiere un aspecto fundamental: que los datos contengan al menos una dimensión geográfica. Usar mapas sirve para contextualizar un problema y para describir un fenómeno en una región determinada y para explicar a la causalidad de un problema en un país o región.

## Cartogramas

Los mapas o cartogramas muestran datos de cantidad asociados a respectivas áreas, mediante la modificación de los tamaños de las unidades de enumeración. La información es presentada mediante la distorsión de las superficies reales, utilizando cada superficie de enumeración como un símbolo proporcional, el cual aumenta o disminuye en función de los valores correspondientes. Un ejemplo podría ser la representación de los países, donde su tamaño en el diagrama dependiera del número de habitantes.



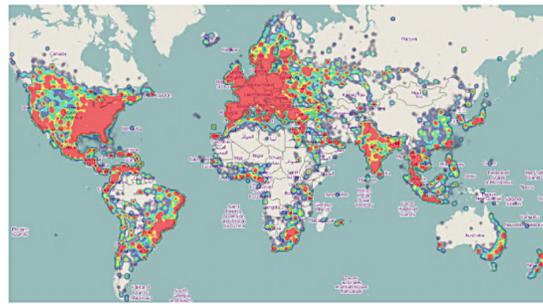
### Mapa coroplético

Un mapa coroplético, mapa coropleto o mapa de coropletas, es un mapa temático en el que las regiones se colorean para mostrar una medida estadística, como puede ser la densidad de población o el ingreso por habitante. Este tipo de mapa facilita la comparación de una medida estadística de una región con la de otra, o muestra la variabilidad de esta para una región dada.



### Mapa calorimétrico

Un mapa de calor es una representación gráfica de los datos, donde los valores individuales contenidos en una matriz se representan como colores y transformaciones fractales. A menudo utilizan un sistema de códigos de colores para representar los valores tomados por una variable en una jerarquía o secuencia.

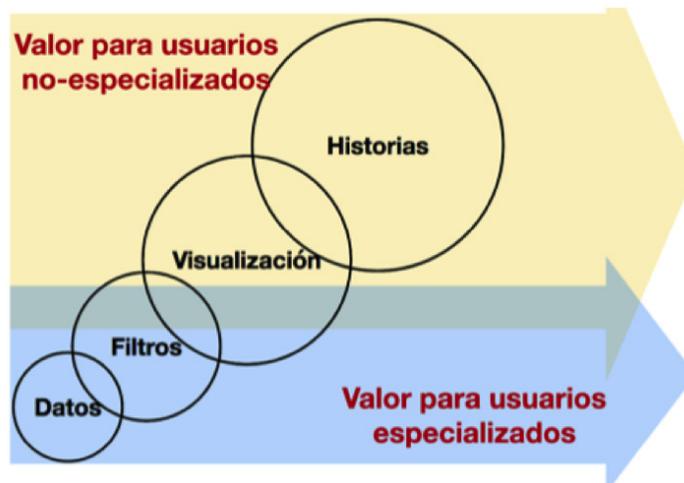


### Formatos diferentes para cada tipo de audiencia

Conocer a tu audiencia va más allá de imaginar a una persona observando tu diseño, conocer a las audiencias que deseas conozcan tu trabajo se traduce en saber cuáles son sus intereses detrás de los datos, que les resulta más relevante en términos informativos, pero también si los usuarios clave interactúan en espacios específicos dentro y fuera de los ambientes digitales.

Las audiencias tienen diversos gustos y necesidades, no se puede comparar a un grupo de investigadores especializados en temas de medio ambiente con un grupo de estudiantes de universidad en la labor de encontrar información para un proyecto escolar. En ambos casos sus necesidades son igualmente legítimas y válidas pero claramente a la hora de interactuar con los datos tendrán estímulos diferentes

Mientras que a algunos usuarios les interesa profundamente interactuar con los datos de origen, o tienen el deseo de realizar filtros específicos de la información; hay otro segmento de usuarios interesados en conocer, por medio de infografías o historias o narrativas digitales, los detalles resultantes del análisis de datos y quedarse con una serie de datos informativos sin mayor profundidad o contexto. En la siguiente gráfica podrás encontrar una explicación visual de ello:



Para conocer más sobre la comunicación visual de los datos se debe comenzar por conocer la diferencia entre las cuatro categorías presentadas anteriormente.

### Datos

Posiblemente la capa con menos usuarios (ya que son usuarios especializados) pero muchas veces son éstos los más comprometidos con el uso intensivo de los datos. Realmente es poco lo que se debe realizar en términos de comunicación para estos usuarios, aunque mantenerlos presentes siempre facilita una conversación sobre un mejor aprovechamiento de la información.

Indicator	TREND	2013 #	2014 #	2015 #	2016 #	2017 #
CPI: annual average inflation (%)	↘	5.23	4.52	0.80	-0.02	-
Credit to Private Sector: % of GDP, end of period	↗	40.42	43.15	44.48	47.20	-
Current Account Balance: % of GDP	↘	-4.94	-4.89	-3.84	-3.55	-
Domestic Pub.Debt: % of GDP-end of period	↗	28.85	29.83	30.89	34.73	-
Fiscal Revenues: % of GDP	↗	14.23	13.96	14.37	14.67	-
GDP: constant prices, year on year growth, %	↗	2.27	3.51	3.57	4.51	-
Gross Fixed Capital Formation: % of GDP	↘	19.53	19.52	19.57	19.18	-
Overall Fiscal Balance: % of GDP	↘	-5.37	-5.61	-5.70	-5.25	-
Public Debt Dollarization: end of period, in %	↗	30.70	36.30	-	-	-
Total Exports (BOP): % of GDP	↗	32.16	33.39	30.95	33.58	-
Total Pub. Debt: % of GDP, end of period	↗	36.00	38.59	41.06	44.79	-
% of salaried employees affiliated w/ social security system, Urban, Lowest Quint, 15-64 age	↗	60.42	54.63	45.25	46.85	-

## Visualización de datos

La visualización de datos tiene un fuerte enfoque en comunicar los resultados del análisis usando la menor cantidad de referencias directas a referencias numéricas sustituyendola por elementos de comparación de total, con base en el número 100 . La visualización de datos se usa frecuentemente para mostrar tendencias, comparar valores o mostrar patrones de comportamiento.



Ejemplo:  
<http://www.100people.org/>

Dentro de esta categoría se considera el uso de infografías, las cuales son representaciones gráficas que combinan diversos elementos tales como imágenes, gráficas y texto para brindar una explicación resumida pero efectiva.

A continuación se identifican 10 recomendaciones para hacer una infografía:

1. Define bien el tema e investiga en profundidad la información que vas a representar
2. Procesa tu información (quita lo que no es necesario).
3. Boceta tu información (haz un diagrama claro de la información).
4. No pierdas de vista lo que a tu audiencia le funciona y los objetivos que quieres lograr.
5. Elige el formato (web o impresión) y el estilo (jovial, serio, divertido, etc.).
6. Con el boceto que tienes haz otro pero como si contaras una historia (principio, nudo y final). Tiene que haber un gancho (algo visual que resalta en la infografía como tipografía, ilustración, diagrama, etc.).
7. Para que tengas mayor claridad revisa los principios básicos del diseño... y la ortografía.
8. Elige un título apropiado, breve y memorable.
9. Pruébala con otras personas y haz modificaciones.
10. Adapta su salida final a los formatos de comunicación (ej. web, redes sociales, impreso, etc.)

## Filtros

En este tipo de entornos se permite que el usuario interactúe con los datos de forma que pueda construir sus propios cruces de información para responder a sus necesidades. En muchos de los casos, la información que se presenta está basada en indicadores o en series de datos en el tiempo de forma que los usuarios ya saben que buscar en este tipo de plataformas

Ejemplo Uruguay:  
<https://databank.bancomundial.org/data/reports.aspx?source=2&country=URY>

Ejemplo Argentina:  
<https://databank.bancomundial.org/data/reports.aspx?source=2&country=ARG>



## Historias

Las historias contadas con los datos surgen a partir de la necesidad de mejorar la interacción y el nivel de comprensión que las audiencias objetivo tienen con los datos. En muchos de los casos las historias basadas en datos interactúan con otros elementos más emocionales en donde su impacto radica en la capacidad de conmovir o conectar con los sentimientos derivado de los datos usados.

Ejemplo  
Ejemplo: <http://especiales.datasketch.co/contratos-colombia/>

Para saber más...  
Accede a este webinar sobre consideraciones y ejemplos sobre narrativas basadas en datos:  
<https://youtu.be/QYIGBOcgsQ>



# Trabajando con geo-referenciación

## Introducción

La georreferenciación es la conversión de localizaciones que los humanos pueden entender en números y códigos que las computadoras pueden entender, como longitud y latitud. Esto permite mapear proyectos, eventos, fenómenos, actividades, hechos, noticias, problemas entre otros.

No es lo mismo decir “estoy en Londres” que decir “estoy en {latitud: -0.81, longitud: 51.745}”

Normalmente, necesitamos georreferenciar cuando queremos que los datos que entienden los humanos (Londres, o un código postal como “12245”) también sean entendidos por las computadoras para que puedan convertir esa localización en un número y, así, ubicarlo en un mapa.

En el ejemplo dado, el término Londres se convierte en un punto en el mapa cuando lo convertimos a latitud y longitud. Por supuesto que Londres es una gran ciudad y no tiene una sola longitud y latitud. Por lo tanto, un polígono representaría de manera más precisa la ciudad. Pero para la mayoría de los ejemplos, con un simple punto basta.

Para georreferenciar, sólo necesitamos una base de datos que nos dé el nombre del lugar y sus correspondientes coordenadas. Hay muchas bases de datos como Open Definition, GeoNames y OpenStreetMap.

## Geo-referenciación

Por suerte, no tenemos que hacer todo el trabajo nosotros. Para eso necesitas saber programar e invertir mucho tiempo.

Existe una amplia variedad de servidores que geocodifican. Te recomendamos usar la base de datos de MapQuest Nominatim que usa bases de datos de mapas. Para encontrar georreferenciación de e calles puedes utilizar Open Street Map. Este servidor te da dos tipos de datos: los que podemos leer los humanos (HTML) y los que pueden leer las máquinas (JSON y XML).

Los datos vectoriales pueden representar tres tipos distintos de entidades: puntos, líneas y polígonos.

- En el caso de los puntos, cada punto tiene coordenadas X-Y que definen su ubicación en el espacio.
- En el caso de las líneas, cada línea tiene un par de coordenadas para cada uno de sus vértices. El ejemplo más sencillo es una recta, que se representaría como la línea que existe entre las coordenadas de su punto de inicio y las de su punto final.
- Los polígonos son datos con coordenadas para cada uno de los vértices que lo componen. Un polígono podría pensarse como el área que se limita por un conjunto de líneas consecutivas que inician y terminan en el mismo punto.

Vale mencionar que una característica importante de los datos vectoriales es que cada una de sus entidades (cada línea, punto o polígono de un archivo) puede tener una tabla de atributos asociados. Esto significa que, por ejemplo, cada sección de un camino puede tener una tabla de atributos con información como “material de que está hecho”, “número de carriles”, “grosor del pavimento”, etc.

Los datos raster, en cambio, son una malla (o dicho técnicamente, una matriz) donde cada celda (o pixel) tiene un tamaño similar y un valor específico. En el caso de las imágenes a color, el raster estaría compuesto por tres matrices sobrepuestas, cada una con celdas que tienen el valor correspondiente a un color primario (enlace a teoría del color). Cuando se utilizan imágenes de satélite u otros sensores (multiespectrales o hiperespectrales) cada archivo puede contener hasta cientos de matrices que representan distintos rangos de lo observado.

Aparte de imágenes, es común tener en archivos raster modelos de elevación digital, que son “imágenes” (una sola matriz) con los valores de elevación promedio del área que corresponde a cada celda.

### **Recursos para aprender más sobre geo-referenciación**

Taller manejo de mapas desde Bases de datos:

<https://youtu.be/z7-hRNDyOoE>

Levantamiento de datos con Kobo:

<https://socialtic.org/blog/guia-kobo-levantamiento-comunitario-de-datos/>

Creación de mapas en QGIS:

<https://es.schoolofdata.org/2019/04/01/creacion-de-mapas-en-qgis/#more-2715>

Haciendo mapas con ArcGIS en línea:

<https://es.schoolofdata.org/2017/10/12/haciendo-mapas-con-arccgis-en-linea/>

Geodatos con Python:

<https://es.schoolofdata.org/2017/11/02/geodatos-con-python/>

### Introducción

Esta unidad desarrolla el concepto de limpieza de datos así como su uso aplicado a diversas circunstancias, ya sea por la falta de estandarización en los campos dentro de una base de datos o por la necesidad de trabajar con un segmento del universo de datos disponibles, la limpieza de datos aporta una serie de conocimientos conceptuales y prácticos dentro del proceso de uso de datos.

### 1) Limpieza y estandarización de datos

La limpieza de datos es el conjunto de técnicas aplicadas a una base de datos con el fin de estandarizar y homologar los términos con los que se describe una categoría, eliminar o corregir errores de escritura o captura, pero también para separar bloques de información más relevantes o menos complejos dentro de una base de datos, con el objetivo de facilitar su procesamiento en etapas posteriores del proceso de datos.

Pocas veces se habla del valor de la limpieza de datos porque en muchas ocasiones es imperceptible su existencia. En momentos en los que cada día más las instituciones se suman a la práctica de abrir datos en formatos abiertos, la expectativa mínima por parte de los usuarios es que esos datos estén diseñados con altos estándares de calidad y por tanto, libres de errores respecto a la calidad de los datos que van a recibir. Sin embargo, no debemos olvidar que las instituciones públicas aún están en medio de procesos transformadores respecto a lo que a tecnologías de gestión documental se refiere y por tanto no toda la información que las instituciones mantienen actualmente se encuentra en óptimas condiciones de ser publicada.

En ocasiones, la limpieza de datos puede entenderse como un proceso para transformar el contenido de una base de datos codificada (es decir que los valores se muestran en forma de códigos para referirse a una zona, un nombre o una categoría) a valores más descriptivos de la información para facilitar su interpretación. Los valores para entender las categorías codificadas normalmente se encuentran en el diccionario de códigos asociado a los metadatos de las bases de datos, de ahí la importancia de conocer previamente el contenido de la base de datos para reducir la incertidumbre al momento de interpretar la información.

En otros escenarios, la limpieza de datos también aporta una capa de conocimiento útil al usuario al comenzar a interactuar con los datos sin indispensablemente estar iniciando un proceso de análisis de la información. El editor de datos de la revista Quartz elaboró una guía sobre puntos a considerar al hacer limpieza de datos. La guía está dividida en cuatro categorías de problemas:

- Problemas que debería resolver tu fuente.
- Cuestiones que deberías resolver tú mismo.
- Problemas que un tercero experto debería ayudarte a solucionar.
- Problemas que un programador debería ayudarte a resolver.

Cada una de estas categorías describe diversos escenarios a encarar durante la limpieza de datos, por ejemplo:

- Valores faltantes.
- Hay datos faltantes que fueron reemplazados con ceros.
- Faltan datos que sabes que deberían estar ahí.
- Filas o valores que están duplicados.
- La ortografía es inconsistente.

La guía completa está disponible en español :

<https://es.schoolofdata.org/guia-quartz/>

Dos consejos esenciales a la hora de limpiar datos son:

- Documentar cada uno de los cambios realizados al limpiar datos (ej: sustituir o corregir el nombre de una calle, separar o unir nombres y apellidos en columnas, etc)
- Guardar versiones del archivo conforme se hacen

## 2) Errores más comunes en limpieza de datos

La ortografía es inconsistente. Los dedazos son una de las maneras más obvias de saber si los datos se compilaron a mano. No sólo te fijas en los nombres de la gente, ese es uno de los sitios donde más difícil es hallar dedazos. En lugar de esto, busca lugares donde los nombres de estados o ciudades no sean consistentes. (“Los Ángeles” como “Los Angeles” es un error muy común, por ejemplo).

Si encuentras errores de este tipo, puedes estar seguro de que los datos fueron compilados y editados a mano y esa es razón suficiente para guardar cierto escepticismo. Los datos editados a mano son los más propensos a fallas. Esto no significa que no deberías usarlos pero entonces deberías corregirlos manualmente o publicarlos como errores en tu reporte.

La herramienta de Open Refine para agrupar texto puede ayudarte a hacer ese proceso sencillo y eficiente al sugerir coincidencias cercanas entre valores inconsistentes en una columna (por ejemplo, igualando Los Angeles con Los Ángeles). Cérciorate, no obstante, de documentar los cambios que haces, de modo que garantices un buen origen de los datos.

### Formatos de fecha inconsistentes:

¿Cuál de estas fechas es septiembre?: 10/9/15 o 9/10/15

Si la primera fue escrita por un latinoamericano o europeo y la segunda por un estadounidense, entonces ambas lo son. Si no conoces la historia de los datos no puedes estar seguro. Averigua de dónde provinieron tus datos y cerciórate de que fue creada por personas del mismo continente.

### Texto que fue convertido a números

No todos los numerales son números. Por ejemplo, la oficina de censos estadounidense usa códigos numéricos para identificar cada sitio en el país. Estos códigos numéricos son de distintas longitudes. No obstante, no son números. Por ejemplo, 037 es el código para el condado de Los Ángeles, no el número 37. Los numerales 37 son, sin embargo, un código de censo válido: para Carolina del Norte. Excel y otras hojas de cálculo con frecuencia cometen el error de asumir que los numerales

son números y les quitarán el cero que les antecede. Esto puede causar toda clase de problemas si tratas de convertir el archivo a otra extensión o cruzarlo con otro set de datos. Cuidado con los datos en los que esto ha ocurrido antes de dártelos a ti.

### Números que fueron guardados como texto

Cuando trabajas con hojas de cálculo, los números pueden ser almacenados como texto con un formato no deseado.

Lo anterior ocurre con frecuencia cuando una hoja de cálculo está optimizada para presentar datos más que para ser reutilizada. Por ejemplo, en vez de representar un millón de dólares con el número "1000000", una celda puede contener el string (secuencia que no es un número) "1,000,000" o "1 000 000" o "USD 1,000,000" con el formato de comas, unidades y espacios ingresados como caracteres. Excel se puede hacer cargo de ciertos casos simples con sus funciones integradas, pero usualmente necesitarás usar fórmulas para sacar o dejar puros caracteres hasta que las celdas estén lo suficientemente limpias para ser reconocidas como números. Una buena práctica es almacenar números sin formato e incluir información contextual en los nombres de columnas o metadatos

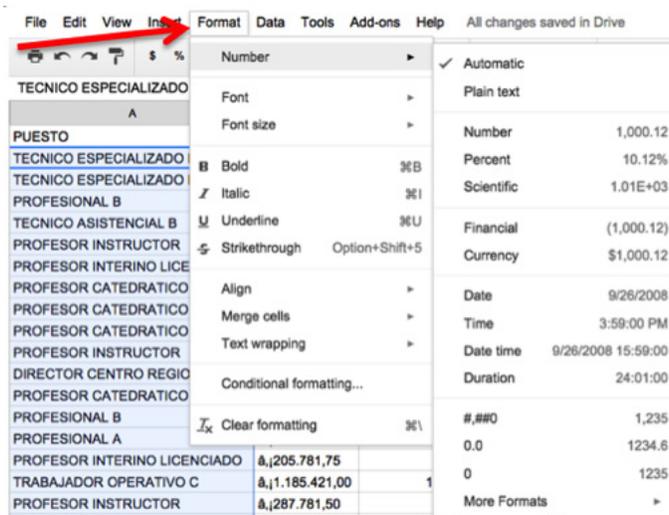
### 3) Limpieza - Modificación del formato de los datos

En la base de datos del ejemplo tenemos solamente cuatro variables que deberían tener los siguientes formatos:

- Puesto: texto
- Salario: numérica
- Jornada: numérica
- Años de servicio: numérica.

Para verificar el formato de los datos nos posicionamos en la primera observación de cada columna y con ayuda de CTRL+SHIFT+flecha hacia abajo, seleccionamos toda la columna.

En la barra superior buscamos la opción que diga Format (Formato) y damos click en la primera opción que dice Number (Número). Al hacer esto se despliega una lista con los diferentes posibles formatos de la variable, y seleccionamos el formato correcto para cada una de las variables.



**TIP:** las variables de texto se alinean a la izquierda y las numéricas a la derecha, por lo que si vemos un número alineado a la izquierda es una mala señal. Otra forma de identificar si el formato numérico está correcto es realizando una operación sencilla, por ejemplo una suma.

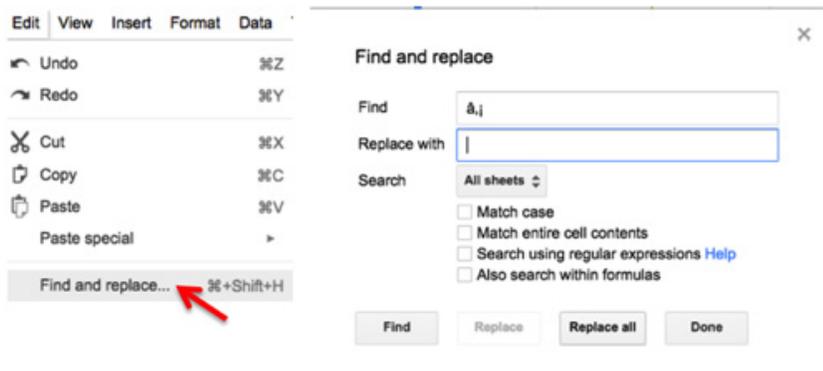
Si da error, posiblemente el formato no sea el adecuado. En la base de datos de salarios, si bien cambiamos el formato de la columna salario a numérico, nos damos cuenta que los números siguen alineados a la derecha. Esto se debe a que además de números tenemos caracteres de texto (que se repiten en todas las celdas), por lo que tenemos que eliminarlos.

#### 4) Limpieza - Buscar y reemplazar

Para limpiar errores usamos la función buscar y reemplazar. Primero seleccionamos con CTRL+C los caracteres que queremos eliminar.



Luego damos clic en Edit (Editar) y buscamos la opción que diga Find (Buscar) and Replace (Reemplazar). En el recuadro que se despliega pegamos los caracteres que queremos reemplazar en Find y en Replace with dejamos el recuadro en blanco, ya que queremos eliminar dichos caracteres.



No obstante, si a pesar de que quitamos los caracteres los números siguen alineados a la izquierda, esto se debe a que dependiendo de cómo tenga configurado su Excel, el programa separa miles y decimales con comas y puntos, es decir, la cifra puede aparecer como 5,200.25 o 5.200,25. Si el Excel está en español normalmente se utiliza la coma para separar decimales y el punto para separar miles; caso contrario si el Excel está en inglés. Es decir ¡cuidado con los puntos y comas cuando trabaja con números!

Para solucionar el problema usamos nuevamente la función Find and Replace. Es muy importante usar la lógica para identificar qué es el valor que queremos sustituir primero. En este caso ponemos en Find "." (un punto) y Replace lo dejamos en blanco, ya que lo que queremos eliminar

son los puntos. Damos clic en Replace all (Reemplazar todos). Luego hacemos lo mismo pero sustituyendo las comas por puntos. Al hacer esto solucionamos el problema y la variable queda guardada como numérica.

<b>SALARIO</b>	
1,722,871.00	
1,631,318.00	
2,363,411.00	
1,200,646.00	

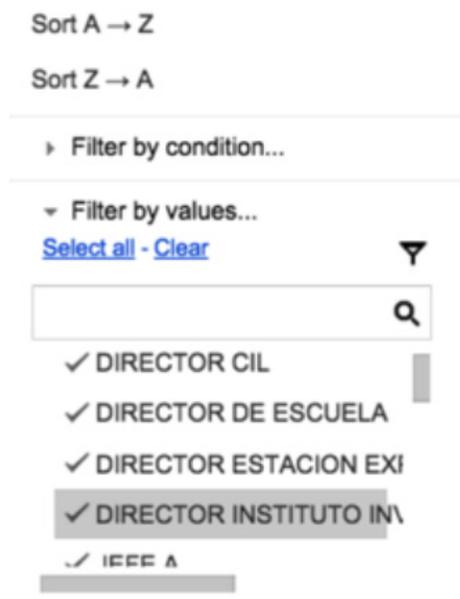
## 5) Limpieza - Filtrar y ordenar datos

Otra forma de identificar posibles errores es por medio de filtros. Los filtros permiten establecer criterios específicos sobre la información de interés que estamos buscando, por ejemplo ordenar los datos de mayor a menor, identificar las categorías de las variables o seleccionar solamente unas de estas categorías.

Para aplicar un filtro, posíciónate sobre la primera fila en el encabezado de la variable que desee filtrar, y en la parte superior seleccione la figura que parece un embudo. Luego de clic en el triángulo que aparece debajo de la celda.



En este ejemplo supongamos que queremos ver cuáles son las diferentes categorías de puestos, por lo que damos clic en el triángulo para ver las opciones de la variable. Al hacer esto nos damos cuenta que hay diferentes categorías para profesores, directores, técnicos, etc, por lo que podríamos tratar de unificar las categorías para tener menos. Esto veremos cómo hacerlo con el programa Open Refine (ver <http://openrefine.org>).



También podríamos querer ordenar de mayor a menor la variable de salario para ver si hay valores muy bajos. Para hacer esto solamente seleccionamos la opción en el recuadro que diga Sort A!Z (menor a mayor) o Sort Z!A (mayor a menor).

Conoce más sobre OpenRefine en tutorial “Encuentra errores y límpialos de manera fácil con OpenRefine” disponible en:

<https://es.schoolofdata.org/2018/05/03/algoritmos-y-clusters-encuentra-errores-y-limpialos-de-manera-facil-con-openrefine/>

## Guías y tutoriales prácticos

Guía para la identificación y uso de entidades interoperables

Subsecretaría de Información Pública y Gobierno Abierto de la Secretaría de Modernización han elaborado esta guía que busca ayudar a los organismos a instrumentar la Política de Datos Abiertos. Ver en

<https://datosgobar.github.io/paquete-apertura-datos/guia-interoperables/>

Con el fin de sistematizar los procesos de limpieza de datos para su publicación en función de los estándares planteados por la Subsecretaría de Información Pública y Gobierno Abierto de la Secretaría de Modernización se ha desarrollado una aplicación en Python. Data Cleaner le permite a los usuarios que pretendan publicar datos abiertos cumpliendo con reglas que siguen las recomendaciones de la Guía para la identificación y uso de entidades interoperables.

Los siguientes tutoriales te podrán ayudar para guiarte en caso de que estés iniciando en el manejo de datos o tengas conocimientos de programación.

**Cómo limpiar datos con Excel (cuatro funciones que cambiarán tu vida)** <https://es.schoolofdata.org/2016/03/21/introduccion-a-limpieza-de-datos-con-excel-cuatro-funciones-que-cambiaran-tu-vida/>

**Algoritmos y clusters: Encuentra errores y límpialos de manera fácil con OpenRefine**

<https://es.schoolofdata.org/2018/05/03/algoritmos-y-clusters-encuentra-errores-y-limpialos-de-manera-facil-con-openrefine/>

**Limpieza de datos con Pandas y Jupyter Notebook de Python**

<https://youtu.be/MqG6z4Bq9F8>

**Limpieza de datos con Python II: manipulaciones, transformaciones, pivotes**

<https://youtu.be/JWXsrSIhmhE>

**Data Cleaner**

<https://data-cleaner.readthedocs.io/en/latest/>

### Introducción

Analizar datos comienza por la formulación de preguntas de diversos niveles de complejidad en la que los datos juegan un papel relevante para la revelación de la respuesta. Esta unidad se enfoca en desmitificar la capa de complejidad que esta etapa arrastra para colocarlo en la correcta dimensión sobre lo que realmente significa analizar e interpretar bases de datos.

Esta unidad a su vez contiene recursos adicionales para aquellos usuarios que deseen profundizar su conocimiento en otros ambientes de trabajo con los mismos objetivos de interpretación de información.

### 1) Análisis de Datos

El análisis de datos posiblemente sea una de las etapas más esperadas de todo el proceso de trabajo con datos; ya que dará respuesta a incógnitas que se plantean incluso desde la motivación inicial de buscar y de usar datos por parte de los usuarios. También, el análisis de datos es observado con cierto grado de complejidad por su natural cercanía al uso de métodos estadísticos para la interpretación del contenido de la base de datos, pero en la práctica no es complejo.

Hacer análisis de datos va más allá de realizar fórmulas matemáticas. Conlleva todo un proceso de pensamiento en el cual se deben formular “preguntas” que deban ser respondidas mediante los datos, de tal forma que se pueda llegar a una correcta interpretación de los datos en el marco de una necesidad específica de información.

El análisis de datos puede realizarse de diversas formas y conlleva ciertos retos pero en general una práctica generalizada es el uso de la estadística descriptiva como método para analizar un conjunto de datos de tal forma que se pueda interpretar lo que la base de datos “tiene que decir”. Este método es muy usado ya que permite mediante el uso de fórmulas básicas conocer el comportamiento de la muestra determinada en la base de datos que se trabaja. Algunos de los cálculos más usados en la estadística descriptiva son:

- El promedio
- La moda
- La mediana
- La variación porcentual
- La desviación estándar
- El uso de tasas

Sin lugar a dudas, también es importante mencionar que cada rama temática requiere de un método específico de interpretación de los datos, por ejemplo al trabajar con datos que contengan información de orden geoespacial se requiere de otro tipo de herramientas y conocimientos, lo mismo sucede con información profundamente ligada al sector económico que igualmente requiere de conocimientos previos para interpretar los datos. Sin duda, es necesario decir que el aprendizaje continuo e

influenciado de varias ramas de conocimiento proveerá a quien requiera de analizar datos de mejores herramientas y métodos para realizar análisis más complejos y diversos con los conjuntos de datos.

## 2) Análisis de datos en hojas de cálculo

Es común analizar datos mediante el uso de hojas de cálculo electrónicas. Su facilidad de trabajo aporta diversos beneficios a usuarios interesados en el manejo de datos. Ahora cada día más los programas de hoja de cálculo permiten hacer análisis más complejos mediante el uso de fórmulas simples o el uso de tablas dinámicas. Estas funcionalidades tienen sus particularidades respecto a cuándo y cómo utilizar e interpretar los resultados obtenidos. A continuación veremos una serie de explicaciones:

## 3) Uso de fórmulas simples

Los programas hoja de cálculo fueron diseñados para que el usuario pueda aplicar una serie de fórmulas para realizar cálculos específicos a los datos con los que trabaja. Las hojas de cálculo contienen diversas librerías de fórmulas y funciones de cálculo, aplicables a diversos tipos de análisis de la información, como lo son el cálculo financiero, estadístico, contable, matemático, entre otras librerías. Gracias al uso de las fórmulas y funciones, el manejo y análisis de grandes cantidades de datos se simplifica al poder realizar diferentes procesos con un par de clics.

Para usar una fórmula o función es necesario conocer la sintaxis con la cual funcionan. Aquí un ejemplo simple:

=nombre\_de\_la\_formula(Argumento1;rango;(fórmulaanidada)\*operadores "constantes")

### Simbolo =

Toda fórmula en una hoja de cálculo inicia con el símbolo "=" que es la manera en que el programa interpreta que se comenzará una operación. Es posible realizar cálculos directos usando operadores matemáticos como:

- + para sumar
- para restar
- \* para multiplicar
- / para dividir
- % para porcentaje
- ^ para potencias

Por ejemplo =5\*4 (5 x 4)

Las diversas librerías que contienen las hojas de cálculo facilitan realizar todo tipo de operaciones. Es normal que una vez que se coloca el símbolo "=" se activa un asistente para ayudar al usuario a encontrar la función que necesita.

=suma(value1, value2,etc...)

### **Argumentos**

Los argumentos son las instrucciones que una fórmula o función requieren para realizar la operación solicitada. Algunas veces los argumentos pueden ser la selección de un rango de celdas o el indicativo de comparar un valor con otro. Los argumentos siempre se separan mediante el uso del símbolo “;” aunque en algunas ocasiones por la configuración de idioma de la computadora del usuario se usa el símbolo “,”

### **Rangos de celdas**

Los rangos de celdas son el conjunto de filas y columnas seleccionadas a las cuales realizar un cálculo. Por ejemplo en la fórmula Suma se puede indicar la celda de inicio y la celda final de donde se encuentra la información a sumar.

```
=sum(a1:a9)
```

En este caso, la hoja de cálculo interpreta que debe sumar los valores desde la celda A1 hasta la celda A9. Vale la pena destacar que en los rangos de celda siempre se separa el punto de inicio y final mediante el uso del símbolo “:”

### **Formulas anidadas**

En los diversos sistemas hoja de cálculo es posible realizar acciones complejas mediante el uso de diversas fórmulas. Esta acción es posible haciendo formulas anidadas, esto quiere decir que se puede incrustar una fórmula dentro de otra fórmula por ejemplo:

```
=PROMEDIO(A14:A28)*(A14*5)
```

Esta fórmula anidada contiene dos operaciones que se resuelven desde el paréntesis interior hacia el paréntesis exterior. En este caso, la fórmula se resuelve multiplicando el valor de A 14 por 5, y posteriormente multiplicando el resultado del promedio de las celdas A14 hasta la A28

Para quien comienza a usar una hoja de cálculo es posible que conocer las diversas fórmulas pueda parecer una misión complicada. Es importante saber que existen diversas fórmulas que sirven para diversos tipos de trabajo. Estas fórmulas y funciones están distribuidas en las diversas librerías de trabajo en las que el formato del uso de fórmulas y los argumentos se presentan de formas similares de forma permanente.

## **4) Análisis de datos usando Tablas Dinámicas**

Las tablas dinámicas o conocidas también como pivot tables, son una forma interactiva de analizar e interpretar la información. Su uso está enfocado a permitir que el usuario pueda ordenar y visualizar grandes volúmenes de datos de diversas formas (información agregada o desagregada) para facilitar su entendimiento.

Cuando se trabaja con conjuntos de datos de muchas filas o columnas, realizar algunas operaciones de análisis podría convertirse en una tarea complicada, esto por la dificultad que representa desplazarse a lo largo de todos los datos. Las tablas dinámicas por su capacidad propia de agregar y desagregar la información permiten que, mediante unos cuantos clicks, la información se consolide en rangos específicos de datos y a su vez se le puedan realizar algunas acciones aritméticas simples entre otras.

Para usar una tabla dinámica es necesario considerar algunos aspectos importantes:

- Es necesario tener una base de datos consistente y limpia, esto quiere decir que la información debe estar libre de errores o símbolos que impidan su procesamiento.
- Algunas computadoras, por lo limitado de sus recursos en memoria RAM y procesador, pueden llegar a presentar algunos errores o a tardar demasiado tiempo.

Las tablas dinámicas están especialmente diseñadas para los usos siguientes:

- Obtener perspectiva sobre grandes cantidades de datos en varias formas.
- Calcular totales y subtotales, resumir datos por categorías y subcategorías.
- Expandir y colapsar niveles de datos para concentrarse en los resultados y explorar en profundidad los detalles desde los datos de resumen para las áreas de más interés.
- Desplazar campos de datos a columnas o filas para ver diferentes resúmenes de los datos de origen.
- Filtrar, ordenar, agrupar y aplicar formato condicional al conjunto de datos.

Para conocer más sobre cómo hacer y usar tablas dinámicas puedes ver el siguiente webinar: <https://youtu.be/UTbnf0bAspQ>

## 5) Porcentajes

Si tenemos un total de N elementos y queremos calcular qué porcentaje representan k elementos de ese total aplicamos la fórmula:

$$\% = \left( \frac{k}{N} \right) * 100$$

Por ejemplo: Un total de 700 personas fallecieron en 2013 por accidentes de tránsito. De estos 170 se debieron a accidentes en motocicleta. ¿Qué porcentaje representan los accidentes en motocicleta del total?

Siguiendo con el ejemplo anterior, supongamos que para el 2014 los accidentes de tránsito aumentaron a 745. ¿Cuál fue el cambio porcentual en el número de accidentes? Usamos la fórmula:

$$\text{Cambio porcentual} = \left( \frac{\text{Valor nuevo}}{\text{Valor anterior}} - 1 \right) * 100$$

$$\text{Cambio porcentual} = \left( \frac{745}{700} - 1 \right) * 100 = 6,4\%$$

En ese mismo año (2014), se registraron 224 accidentes de motocicletas, lo cual representa un 30% de los accidentes de tránsito de ese año. En comparación al 2013 los accidentes de motocicletas aumentaron en 5.8 puntos porcentuales (30-24,2). Por su parte el cambio porcentual en este

tipo de accidentes fue de un 31,7% .

Confundir estos dos conceptos es un error común. Así por ejemplo si la tasa de interés pasó de 3% a 4%, lo correcto sería decir que aumentó un punto porcentual o un 33,3%, mientras que lo incorrecto sería decir que aumentó un 1%.

## 6) Ajuste por inflación

Cuando comparamos cantidades monetarias de diferentes periodos es necesario ajustarlos por la inflación. Esto se debe a que por ejemplo \$1 en el 2015 vale mucho menos que \$1 hace 10 años. Cuando hacemos la comparación sin ajustar por inflación estamos utilizando los montos nominales , mientras que si hacemos el ajuste serían montos reales.

¿Por qué es importante? Supongamos que estamos haciendo un trabajo sobre el gasto del gobierno en un programa específico. En 2010 se gastaron \$21,6 millones mientras que en el 2014 el gasto aumentó a \$27,5 millones. Si hacemos la comparación en términos nominales obtenemos que el gasto en el programa aumentó en un 27%. No obstante realizar esta comparación no refleja el cambio en el valor de la moneda, es decir, la cantidad de cosas que podíamos comprar en 2010 con ese monto versus las que podemos comprar en el 2014. Es por eso que es recomendable hacer el ajuste por inflación.

Para ello utilizamos normalmente el Índice de Precios al Consumidor que calculan los Institutos Nacionales de Estadísticas. Este indicador permite medir la evolución de los precios de un conjunto de bienes y servicios específicos. Este índice tiene un periodo base igual a 100 y varía de periodo a periodo.

Siguiendo el ejemplo anterior, si quisiéramos convertir los montos del programa del gobierno a pesos constantes (o reales). Supongamos que el IPC para el 2014 fue de 171,2 y el de 2010 de 141 (la base utilizada en este caso es la de 2006). Procedemos de la siguiente manera:

a) Convertimos cada uno de los montos a pesos (moneda) del año base:

$$\left( \frac{\text{monto}}{IPC} \right) * 100$$

En el ejemplo sería, para 2014  $(27,5/171,2)*100 = \text{¢}16,06$  millones de pesos de 2006. Mientras que para el 2010  $(21,6/141)*100 = \text{¢}15,31$  millones de pesos de 2006.

b) Una vez que tenemos los montos en términos reales calculamos la variación porcentual.  $((16,06/15,31)-1)*100 = 4,89\%$

Es decir el gasto en el programa solamente aumentó un 4,9% en términos reales entre 2010 y 2014, mientras que si no hubiéramos hecho el ajuste por inflación la variación hubiera sido de 27%.

¿Qué pasa si queremos expresar los montos no en pesos de 2006 sino en pesos de 2014? Para ello creamos una nueva base. En este caso si nuestra

nueva base es la de 2014 lo que hacemos es dividir los diferentes IPC de los periodos entre el IPC del nuevo año que queremos tomar como base. De esta manera en el ejemplo el nuevo IPC para 2014 sería  $171,2/171,2=1$  mientras que el 2010 sería igual a  $141/171,2=0,8205$ .

Es importante notar que ahora el IPC no está en una base 100, sino en base decimal, por lo que para calcular los montos en pesos de 2014 únicamente tenemos que dividir el monto entre el índice sin multiplicar por 100.

Así tendríamos que para 2014 el monto sería  $\$27,5$  millones mientras que el 2010 sería  $21,6/0,8205= 26,32$  millones. Si ahora calculamos la variación porcentual obtenemos que es de 4,8%, igual que al calcularla de la otra manera.

## 7) Comparar “manzanas” con “manzanas”

Supongamos que tenemos que el gasto para el 2014 del municipio de la ciudad capital San José fue de \$60.000 millones y el de Garabito (una comunidad rural) fue de \$5000 millones. Si decimos que la municipalidad de San José gastó 12 veces más que la de Garabito, si bien el dato es numéricamente correcto la comparación es engañosa. ¿Por qué? Mientras la ciudad central de San José tiene 333.980 habitantes la de Garabito tiene 22.767 habitantes.

Así una mejor forma de comparar es calcular el gasto per cápita dividiendo el gasto total entre el número de habitantes. Al hacer dicho cálculo obtenemos que el gasto per cápita de la municipalidad de San José fue de 179 mil pesos, mientras que el de la municipalidad de Garabito fue de 219 mil pesos. Es decir el municipio rural tiene un mayor gasto por habitante que el de la ciudad capital, lo cual da un nuevo ángulo a nuestra historia y hace la comparación más adecuada.

Otras formas útiles de hacer este tipo de comparaciones es calcular Tasas. Por ejemplo la tasa de homicidios por cada 1000 o 100 mil habitantes. Esta se calcula:

$$\left( \frac{\text{Número de ocurrencias del evento (homicidios, accidentes, nacimientos, etc)}}{\text{Total de la población}} \right) * 100000$$

Esto me daría la tasa de homicidios (por ejemplo) por cada 100 mil habitantes en una localidad determinada.

Ya vimos que para hacer comparaciones más precisas es necesario comparar dos cosas de la misma categoría, por ejemplo hacer las comparaciones en términos reales o tomando en cuenta a la población de una localidad.

Un error común cuando se comparan datos de periodos es comparar periodos distintos. Por ejemplo comparar el total de turistas que llegaron a un país durante TODO el 2014 con el total de turistas que han llegado al país de enero a junio del 2015. Esto es incorrecto ya que no estamos comparando los mismos periodos. Lo correcto sería comparar datos del mismo periodo, en este caso el primer semestre de 2014 versus el primer semestre del 2015.

Lo mismo ocurre si por ejemplo no tomamos en cuenta cambios que pueda sufrir una variable entre periodos. Por ejemplo si un Ministerio del Gobierno absorbió funciones de otro entre dos periodos tendría sentido que su presupuesto aumente. Estos cambios también se pueden deber a la forma en la que se miden indicadores como la pobreza, que pueden tener variaciones metodológicas entre periodos lo cual puede hacer los datos no comparables. En estos casos hay que estar pendiente de estos detalles y de ser necesario consultar a la fuente.

## 8) Evitar valores absolutos

“El Ministerio de Seguridad gastó \$200 millones en capacitación a policías” ¿Y? Presentar una cifra así no nos dice mucho, no sabemos si 200 millones es mucho o poco, si es un monto razonable para gastar en ese tipo de proyecto, si es lo mismo que gastó el año pasado.

Es por eso que presentar cifras absolutas puede ser poco efectivo si no las dotamos de contexto. En estos casos es mejor utilizar algún tipo de comparación o contextualizar el dato, por ejemplo decir (si así fuera el caso) que el ministerio gastó tres cuartas partes de su presupuesto en un único programa.

## 9) Unidades de análisis

Supongamos que queremos hacer un artículo sobre el rendimiento académico en los colegios y tenemos dos tipos de datos: la notas por colegios y por estudiantes. Al analizar los datos obtenemos dos tipos de resultados que podrían convertirse en estas afirmaciones:

A: “El rendimiento académico está peor porque 60% de los colegios tuvieron resultados peores que el año anterior”.

B: “El rendimiento académico mejoró este año porque el 80% de los estudiantes obtuvieron mejores calificaciones que el año anterior”.

Las dos afirmaciones son correctas sin embargo, ¿cuál es más adecuada? La diferencia entre ambos casos es que se están utilizando diferentes unidades de análisis, en el ejemplo A se usan los colegios y en el B los estudiantes. En este caso específico sabemos que no necesariamente las escuelas tienen el mismo número de estudiantes por lo que sería más adecuado utilizar como unidad de análisis los estudiantes y no los centros educativos.

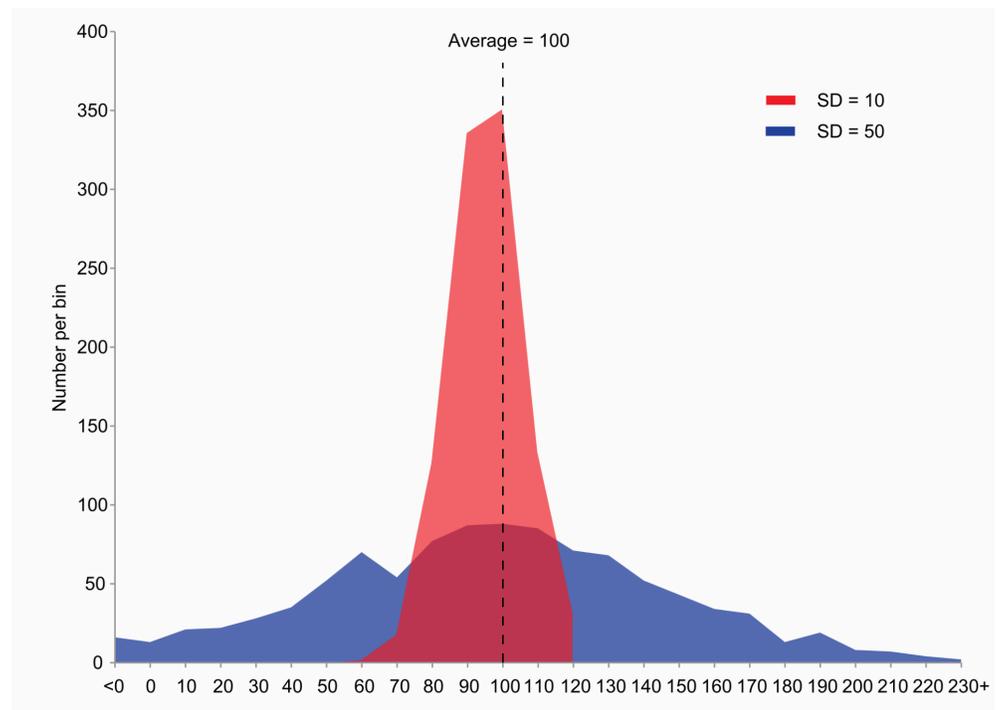
Lo mismo sucedería si queremos analizar si una industria ha empeorado o mejorado en el tiempo: ¿usamos el número de exportaciones, el número de empleos en la industria o ambos?.

Así, la lección es que a la hora de analizar un fenómeno es importante pensar cómo vamos a medirlo y qué variables son más adecuadas para lo que queremos analizar.

## 10) La media (promedio) versus la mediana

La media (o promedio) y la mediana son medidas de tendencia central, que nos permiten ubicar ciertos puntos de la distribución. Para calcular un promedio simple sumamos las observaciones y las dividimos entre el total, mientras que la mediana mide el centro de la distribución, es decir, 50% de las observaciones se encuentran por debajo de la mediana y 50% por encima de ella.

¿Cuál es la medida más recomendable para utilizar?  
Depende.



El promedio se afecta por valores extremos y por la distribución de los datos, por lo que no es una medida muy adecuada en casos con valores de alta variabilidad en los rangos extremos. Por el contrario la mediana es menos sensible a la dispersión de la distribución y no se afecta tanto por los valores extremos. No obstante esto puede ser una desventaja si los valores extremos son de interés para nuestro análisis.

Una buena forma de abordar el problema es calcular ambas medidas y además reportar medidas de dispersión como la desviación estándar o el rango de la distribución. Esto nos lleva al otro punto.

# Introducción al Periodismo de Datos

## Introducción

En esta unidad se abordarán características que han marcado el periodismo de datos latinoamericano en años recientes. De manera consciente, se evitan las definiciones y las teorías sobre el periodismo de datos, puesto que a medida que avanza la tecnología, la apertura de datos y las capacidades de quienes hacen periodismo, la práctica, herramientas y metodologías del uso de datos en el periodismo es una dinámica evolutiva.

Hay que recordar que, tal como inicia El Manual de Periodismo de Datos 1.0 mencionando:

Tanto “datos” como “periodismo” son términos problemáticos. Algunos creen que “datos” es cualquier colección de cifras, por lo general reunidas en una hoja de cálculo. Hace 20 años, esos eran prácticamente los únicos datos que manejaban los periodistas. Pero ahora vivimos en un mundo digital, un mundo en el que casi cualquier hecho puede ser (y casi todo es) descrito con números.

Tal como se ha mostrado a lo largo de este curso, a medida que existe mayor disponibilidad y apertura de datos, sea el que sea su formato, el potencial uso de los mismos es incremental. Este potencial ha dinamizado el periodismo de manera vigorosa pues hoy en día se puede tener la capacidad para generar datos en campo a través de teléfonos celulares, analizar conversaciones masivas en redes sociales, extraer documentos en PDF a formatos abiertos, consultar miles de sitios web de manera automatizada, programar scripts de análisis de bases de datos masivas, organizar información con base en reconocimiento automatizado de texto e inclusive aplicar algoritmos para detectar características específicas de datos (ej. patrones, anomalías, redundancias, etc.) con tal de fortalecer la labor periodística. Este reportaje ayuda a tener una visión general de lo que implica hacer periodismo de datos.

<https://www.youtube.com/watch?v=1qxxtLRYK4k>

## Generación y obtención de datos

A pesar de que las fuentes tradicionales de datos —tales como las fuentes directas o indirectas, la consulta a archivos o hemerotecas y el acceso a la información pública— siguen siendo predominantes en el periodismo; a medida que la agenda de apertura de datos avanza en un país se aumentan las posibilidades de basar indicios e investigaciones periodísticas en datos abiertos. Países con amplios data-sets abiertos se vuelven fuentes no sólo de consultas específicas por parte de periodistas, sino espacios en donde periodistas pueden “entrevistar los datos” (es decir, generar preguntas que desaten análisis de datos) para identificar indicios periodísticos.

En contextos con ausencia de apertura de datos públicos, las labores de datos suelen requerir una combinación más intensiva de esfuerzos. Ante la falta de datos se requiere que periodistas generen datos desde su fuente inmediata. Herramientas para elaborar sondeos, encuestas y documentación in-situ de manera estructurada fortalecen el levantamiento

de datos utilizando formularios y aplicaciones especializadas que funcionan desde teléfonos móviles, que combinan información documentada en formularios con evidencia complementaria como lo son fotografías, audios de entrevistas, videos y datos de georreferenciación.

Así mismo, cuando se requiere extraer información desde espacios digitales, se puede acceder a servicios de consulta de datos en tiempo real, como lo pueden ser APIs, para generar llamados, consultas o descargas automatizadas de la información que se esté investigando. Tal es el caso de los análisis de conversaciones en redes sociales o de la elaboración de programas que generan consultas continuas a sitios web que leen y guardan dicha información de manera estructurada en una base de datos.

También, para acceder formal (ei. acceso a la información) o informalmente (ei. filtraciones) a la información pública se requiere muchas veces transformar los documentos recibidos en formatos cerrados (ej. papel, imágenes o PDFs) a formatos abiertos. Estas necesidades técnicas han hecho que el periodismo de datos requiera un componente multidisciplinario en donde periodistas se alíen con especialistas con perfiles de ciencia de datos y programación de software.

Ejemplos:

- Dos años de las escuchas de Nissman - <https://www.lanacion.com.ar/politica/dos-anos-de-analisis-de-las-escuchas-de-nisman-nid1976325>
- VenData - <https://vendata.org/site/>
- Cuentas Juradas - <https://cuentasjuradas.ojo-publico.com/>

### **Datos como indicio**

A medida que el periodismo utiliza más los datos, el sentido u olfato que le habilita a encontrar aquellas pistas, rumbos o indicios donde puede haber grandes historias se va perfeccionando a través de la exploración de datos. Así como tradicionalmente se pueden indagar fuentes, archivos o vincular información documental, el periodismo de datos busca identificar el punto de partida de posibles historias en los datos mismos, ya sea sistematizando información que potencialmente al estar en bases de datos pueda validar hipótesis, como analizando datos abiertos con tal de identificar patrones o anomalías que puedan desatar revelaciones que el olfato análogo jamás brindaría. En este enfoque de uso de datos, se vale utilizar la expresión "entrevistar a una base de datos" al establecer hipótesis que son probadas a través de distintos análisis de datos.

Al identificar indicios basados en datos, el esfuerzo no queda ahí, pues es el punto de partida de distintas metodologías periodísticas tales como la documentación, el reporte, la verificación de información, el acercamiento a fuentes relacionadas el tema y la documentación de evidencia periodística de sustento.

Ejemplos:

Mi Edificio - <https://miedificio.contralacorrupcion.mx>

La Estafa Maestra - <https://www.animalpolitico.com/estafa-maestra>

## Datos como argumentos

El rol tradicional de los datos en el periodismo es como argumento. A través de los años, el sustento de afirmaciones e historias con "datos duros" en un reportaje periodístico ha evolucionado para dejar de ser sólo una cifra y convertirse en un análisis cuantitativo que sustenta testimonios individuales, opiniones de especialistas y análisis contextuales de la materia que se esté investigando. Son justo estos argumentos basados en hechos y análisis comprobables los que hacen que una investigación tenga mayor legitimidad y profundidad pues puede llegar a contar con diversos elementos (cualitativos y cuantitativos) que permiten que personas lectoras se enganchen, comprendan y se convenzan de lo planteado en la historia.

Ejemplos:

- Sudestada - <https://www.sudestada.com.uy/10887/Datos>
- SOS Riachuelo - <https://www.lanacion.com.ar/sociedad/sos-riachuelo-nid1937766>
- Geografía del Delito - <https://www.eltiempo.com/multimedia/especiales/delitos-homicidios-inseguridad-y-crmenes-en-colombia/15207495/1>

## Datos para la comprensión de fenómenos

Otro enfoque del periodismo de datos es aquél que busca aportar una explicación amplia y cuantitativa de algún fenómeno, tema de interés o coyuntura. Muchas veces, este tipo de investigaciones están basadas en datos públicos disponibles y buscan generar indicios periodísticos, dudas razonables bajo coyunturas, visibilizar inconsistencias y hasta cuestionar las narrativas comúnmente expresadas sin el sustento cuantitativo apropiado. Estos enfoques periodísticos suelen estar dirigidos a públicos especializados en las temáticas que abordan y que tienen el conocimiento suficiente para dar sentido y contexto a los datos, análisis e inclusive espacios interactivos basados en datos que publican.

Ejemplos:

- La Data - <http://www.ladata.mx>
- Serendipia Data - <https://serendipia.digital/2019/03/tomas-clandestinas-aumentaron-en-enero-a-pesar-del-cierre-de-ductos/>

## Datos como plataforma

Equipos periodísticos que de manera sistemática generan y obtienen datos para sus propios proyectos han optado por realizar proyectos en donde no sólo se difunden sus historias sino todos los conjuntos de datos que hayan logrado acumular sobre un tema. En estos casos, generalmente las compilaciones de datos publicados son liberadas con reportajes exclusivos basados en estos datos pero, dada la amplitud y cantidad de los mismos, dejan espacios para que cualquier persona los consulte y utilice para sus propios fines.

Muchos de estos esfuerzos de periodismo de datos implican diversas tareas de acceso a la información, extracción, limpieza y sistematización de datos detrás de lo que una persona usuaria puede llegar a ver o descargar en los sitios

correspondientes a estos proyectos. Esta lógica no sólo refuerza las acciones de apertura de datos sino que también habilita el periodismo u otros proyectos basados en datos de manera colaborativa. E inclusive, puede ser la base de consulta estática (a través de descarga de archivos o consulta personal) o dinámica (a través de consultas automatizadas al API del proyecto) de servicios y proyectos de terceros.

Ejemplos:

- Río Abierto - <https://rioabiertodatos.ladiaria.com.uy/>
- Quien Es Quien Wiki - <https://www.quienesquien.wiki>
- Medicamentalia - <https://medicamentalia.org>

### Datos como verificadores

A medida que el periodismo se diversifica, atiende necesidades características de las dinámicas comunicacionales del siglo 21. Un claro ejemplo de ello ha sido la verificación periodística de hechos, declaraciones y documentos públicos realizado por "fact-checkers" (eí. verificadores de hechos) tanto a nivel documental como en tiempo real. Las técnicas de verificación siempre requieren contar con consulta especializada, profunda y ágil a fuentes de datos que permitan a los equipos periodísticos contrastar los hechos, dichos y noticias expresados públicamente con datos que permitan evaluar un juicio de veracidad sobre éstos. Entre más sistematizada esté la información para su consulta, más posibilita que los equipos de verificación lleguen a los datos correspondientes para realizar los análisis ad hoc a la información para verificar.

Actualmente, esfuerzos de combate a la desinformación (también conocida como "fake news") replican técnicas de verificación de hecho y discurso para identificar noticias falsas.

Ejemplos:

- Chequeado - <https://chequeado.com>
- UY Check - <http://uycheck.com/>
- Derribando Mitos - <http://derribandomitos.com/>
- El Sabueso - <https://www.animalpolitico.com/sabueso>
- Ojo Biónico - <https://ojo-publico.com/ojobionico>

### Narrativas basadas en datos

A medida que los equipos periodísticos se internalizan más en el uso de datos como insumo de su labor, se ha identificado que la visualización de datos va más allá de simplemente insertar gráficas en un texto periodístico. Tal como se menciona en la sección sobre visualización de datos de esta unidad, las gráficas llegan a ser elementos clásicos para representar el comportamiento o desenlace de un análisis de ciertos datos, pero dependiendo del tipo de historia, revelación que haya generado los datos y tipo de audiencia a los que una salida periodística va dirigida, es necesario seleccionar cuidadosamente los elementos que mejor llevarán a la persona usuaria a navegar, consumir e internalizar lo que ahí se plasme.

En espacios digitales, muchos conceptos de diseño de experiencia de usuario (UX) e interfaces (UI) se deben de reflejar en el periodismo basado en datos, pues la combinación de los distintos elementos necesarios para contar una historia deben realizarse de manera intuitiva, comprensible y útil. Este tipo de técnicas requieren que el periodismo comprenda bien a su audiencia para apegarse a sus hábitos de consumo en cuanto a la plataforma y prácticas en las que funciona como persona usuaria de espacios digitales. Así mismo, se deben seguir principios básicos de diseño web, interacción en interfaces digitales y prototipado de productos antes de la publicación final de proyectos periodísticos.

## MÓDULO 4

# Lecturas

Villatoro, D. (2017). Datos con Propósito. Recuperado de <http://tinyurl.com/datosproposito>

Cucho, A. (2014). ¿Cómo empezar mi propia aventura con datos?. Recuperado de <https://es.schoolofdata.org/2014/10/23/como-empezar-mi-propia-aventura-con-datos/>

¡Libera los datos encerrados en archivos PDF! - Blog Abierto al Público: <https://blogs.iadb.org/conocimiento-abierto/es/libera-los-datos-encerrados-en-archivos-pdf/>

Conoce Gmapsdistance: la herramienta que te permite calcular distancias de forma masiva - Blog Abierto al Público: <https://blogs.iadb.org/conocimiento-abierto/es/gmapsdistance-herramienta-calculador-distancias-masiva>

Verhulst, Stefaan G.; Young, Andrew, Toward an Open Data Demand Assessment and Segmentation Methodology: <http://dx.doi.org/10.18235/0001529bank.org/en/essentials.html>

## MÓDULO 5

---

# Uso de datos en mi organización

---

Docente: Dra. Javiera Atenas

TUTORIAS

Ignacio Alfaro  
Sergio Araiza

# **Resumen** **Introducción**

Este módulo tiene como objetivo proporcionar las herramientas y las habilidades necesarias para desarrollar proyectos de datos abiertos a nivel organizacional con un enfoque en la priorización de la apertura de datos, y acceso a la información pública para distintos tipos de usuarios con diversos niveles de alfabetización y capacidades.

La primera unidad apunta a la priorización en la apertura de datos analizando y clasificando a aquellos conjuntos de datos que se pueden considerar de alto valor o de mayor relevancia, para priorizar su publicación.

La segunda unidad provee una introducción a los criterios para la apertura de datos para facilitar que las organizaciones conozcan los métodos para determinar la prioridad de la publicación de datos, por ejemplo, construyendo y co-creando un plan estratégico que contenga una serie de instrucciones, como una guía para la creación de políticas institucionales de datos abiertos.

La tercera unidad es una introducción al desarrollo de planes de publicación de Datos Abiertos con enfoque en los datos prioritarios, en las capacidades organizacionales para abrir datos, en las capacidades del público objetivo, en las metas a alcanzar, y en el valor de los servicios a proveer.

La cuarta unidad es una introducción al diseño centrado en las necesidades de los usuarios con un enfoque en los elementos que pueden beneficiar a los usuarios y a las organizaciones para que el impacto del proyecto sea real.

Finalmente, la quinta unidad es una Introducción a la formulación de planes estratégicos de datos abiertos y al desarrollo de mapas de ruta para determinar qué elementos, recursos, participantes, guías, estructuras, consultas y organización se necesitan para desarrollar un proyecto sustentable que ayude a resolver problemas relacionados con las necesidades de los usuarios.

## **OBJETIVOS DE APRENDIZAJE**

1. Desarrollar proyectos de datos a nivel organizacional
2. Comprender los modelos de apertura de datos centrados en los usuarios
3. Construir protocolos para favorecer la comprensión de la información pública y datos abiertos para distintos grupos de usuarios
4. Diseñar estrategias de capacitación en datos abiertos para distintos grupos de usuarios
5. Definir líneas prioritarias para abrir datos a nivel organizacional

## Priorización en la apertura de datos

### Introducción

Cuando desarrollamos un plan de apertura de datos abiertos, es necesario que quienes publican los datos, sea la academia o el sector público analicen y comprendan cuáles son los conjuntos que podemos considerar de alto valor o de mayor relevancia, para priorizar su publicación. Para poder determinar la relevancia o el valor de los conjuntos de datos a publicar, es necesario considerar ciertos elementos, como el valor de estos para las comunidades de usuarios, para el diseño de políticas o para fomentar la participación pública. También se deben considerar otros elementos que pueden presentarse en casos de contingencia nacional o internacional, como por ejemplo datos sobre emergencias o desastres naturales, epidemias o casos de corrupción, que requieren ser publicados de forma veloz.

### 1.1 Evaluando el valor o relevancia de los datos

Para poder determinar qué datos es necesario priorizar cuando se desarrolla una estrategia de apertura, es importante hacerlo de manera colaborativa e involucrando a una serie de actores, entre los que se incluye la sociedad civil, los usuarios de datos, la comunidad académica y también los usuarios de los servicios (aunque estos no utilicen los datos), analizando diversos factores para determinar si un set de datos es de alto valor.

Los factores a determinar se obtienen mediante consultas a los grupos de usuarios, por ejemplo, la frecuencia de publicación de los conjuntos de datos, y también los niveles de agregación de los datos. Este proceso debe ser evaluado con una cierta periodicidad, consultando a los usuarios si los conjuntos de datos que se han abierto, les han sido útiles y también revisando el impacto que estos han tenido, monitoreando elementos como el número de descargas, su utilización en informes o reportes, si estos han sido solicitados mediante peticiones de acceso a la información pública, si han sido, citados en notas de prensa o si se han utilizado como objeto de estudio.

Los procesos de apertura mediante priorización, permiten a las organizaciones centrarse en las necesidades de los usuarios, lo que puede considerarse como un modelo costo-efectivo, ya que la apertura de datos tiene un alto costo económico, y el priorizar datos puede disminuir el pagar innecesariamente el costo asociado de abrir datos que no son utilizados, ni son de interés social o de investigación.

### 1.2 Elementos a Considerar

Entre los elementos que es recomendable incluir cuando se diseña una estrategia organizacional para priorizar la apertura de datos, es necesario considerar un mecanismo para evaluar la calidad de los datos que se van a abrir y sus componentes éticos. Como por ejemplo, la frecuencia con la cual se piensan recopilar y procesar los datos, y los procesos de gobernanza

que existen para gestionarlos, ya que estos son importante para saber quién administra la publicación de los datos, quién coordina a los equipos responsables de su gestión y quién es responsable de las actualizaciones y de coordinar las relaciones con los usuarios y la sociedad civil.

Es importante también que los usuarios y la sociedad civil sepan quiénes son sus puntos de referencia en las agencias que publican los datos, quién es responsable de administrar y mantener los datos, quién tiene la propiedad de los datos, ya que por ejemplo, esta información es útil para aquellos investigadores que requieren citar un conjunto o set de datos en sus publicaciones.

Al momento de abrir datos, es importante considerar que estos deben ser legibles tanto por máquinas como por personas, y por ende, accesibles para las comunidades de usuarios y la sociedad civil. Por lo tanto, es importante identificar en qué formato se encuentra la información que se planea abrir, el formato en el cuál se va a abrir, y también, el esquema de metadatos o el estándar que será aplicado a los datos (ver módulo II unidad 2: Indicadores y herramientas (ODRA, Open Data Index, Barometer, y unidad 4: Estándares para liberación de datos e Instrumentos de medición).

Otros elementos a considerar en un plan estratégico de priorización de datos, es la vida útil de un set o conjunto de datos, ya que si estos son datos de contingencia y se han abierto de manera apresurada, podrían tener una vida útil corta, por lo que hay que planificar la priorización de apertura, considerando que los datos abiertos, aún publicados en casos de urgencia, deben tener una vida útil que permita que estos datos se utilicen en otros contextos. También, y considerando el elemento contingencia / urgencia, es necesario, calendarizar la frecuencia de la actualización y/o publicación de datos y los mecanismos para comunicar estas actualizaciones o la publicación de nuevos datos a las comunidades de usuarios y a la sociedad civil.

A su vez es necesario establecer procesos de control de calidad de los datos, y tener en cuenta elementos como la fuente primaria de los datos, las metodologías con las cuales se han recopilado, y proveer mecanismos para validarlos, para evitar incertidumbres con relación a su exactitud, credibilidad y validez, proveyendo a las comunidades de usuarios información sobre la proveniencia de los datos.

Finalmente, hay que considerar ante la publicación de un set o conjunto de datos, ciertos elementos éticos y de protección de datos personales. Es necesario evaluar la privacidad de los datos para evitar que estos puedan incluir información personal, o que pueda facilitar la identificación de personas o colectivos, promover la discriminación y/o estigmatización de personas o colectivos vulnerables, por ejemplo, datos de rendimiento educativo de escuelas de zonas con alta población migrante o en situación de precariedad económica.

En algunos casos, los criterios de priorización de la apertura pueden variar o pueden tener ejes o enfoques diferentes, como se puede ver en los ejemplos a continuación

México: En el caso de este país se espera que la priorización de datos de valor apunte lo siguiente “es importante para que los esfuerzos de gobernanza, apertura y publicación estén enfocados en la provisión de datos que sean de utilidad, valor y tengan el potencial de crear impacto para la ciudadanía”.

Perú: En el caso del Perú, su guía apunta a lo siguiente”La Estrategia Nacional de Datos Abiertos Gubernamentales 2017-2021 y el Modelo de Datos Abiertos Gubernamentales del Perú tienen como finalidad promover la apertura de datos de la información de las entidades públicas, la innovación en la generación de valor público con la reutilización de los datos abiertos para la creación de nuevos productos y servicios con el uso de las tecnologías de la información y comunicación para contribuir al desarrollo social y económico, en el marco de un gobierno abierto. Asimismo, buscar promover alianzas público-privadas a través de la apertura y reutilización de la información para mejorar los servicios a los ciudadanos”.

[http://sgp.pcm.gob.pe/wp-content/uploads/2017/01/Guia\\_RapidaDAG\\_v030117.pdf](http://sgp.pcm.gob.pe/wp-content/uploads/2017/01/Guia_RapidaDAG_v030117.pdf)

Otros ejemplos los podemos encontrar en el artículo del blog del BID Apertura y uso de la información legislativa por primera vez en la agenda de condatos

<https://blogs.iadb.org/conocimiento-abierto/es/apertura-y-uso-de-la-informacion-legislativa-por-primera-vez-en-la-agenda-de-condatos/>

### 1.3 Entendiendo el concepto de prioridad.

Para poder entender qué significa el concepto de datos prioritarios, la Carta de Datos Abiertos (Open Data Charter G8) los define como datos que pueden ayudar a mejorar la democracia y fomentar la reutilización innovadora de los datos particulares. Es importante considerar que el priorizar la apertura de un set o conjunto de datos puede tener más o menos relevancia dependiendo de quiénes sean los potenciales usuarios principales de estos datos, por ejemplo:

**Empresas y sector privado:** Para el sector privado, aquellos datos que puedan ser utilizados para proveer servicios, para crear nuevos productos, para insertarse en un mercado tienen alto valor pero en el caso de la sociedad civil, los datos que pueden mejorar la gobernanza son aquellos que pueden considerarse prioritarios. Los emprendedores y empresas, son cruciales para aprovechar el potencial de los datos abiertos, creando nuevos productos y modelos de negocio basados en datos abiertos. Cada vez más empresas tienen un modelo de negocio basado en datos abiertos, ya que estas confían en la disponibilidad y la calidad de los datos y la información pública para poder desarrollar nuevos modelos de negocios, productos y servicios, al poder acceder a una gran cantidad de recursos de forma gratuita para desarrollar, por ejemplo, aplicaciones para dispositivos móviles, o programas computacionales.

**Salud:** Otro escenario interesante a contemplar en relación con la prioridad de datos, tiene que ver con los beneficios que los datos abiertos pueden proveer al sector de la salud. Estos datos pueden ayudar a los hospitales, a las administraciones públicas, a los investigadores y a la ciudadanía, a mejorar los tiempos de atención médica, a desarrollar nuevos protocolos o tratamientos médicos, a prestar atención especializada a grupos vulnerables de la población, y a repartir de mejor manera los fondos públicos para mejorar la calidad de la atención en hospitales públicos y la calidad de la investigación biomédica. Para entender más, les recomendamos mirar este video sobre datos abiertos y salud

**Educación:** Para el sector educativo, los datos abiertos tienen valor en dos dimensiones. Primero, los datos abiertos sirven para observar la calidad de la educación, mediante la evaluación de información que proveen las instituciones educativas, los ministerios de educación y los organismos internacionales, lo que permite que las escuelas que necesitan más recursos económicos puedan acceder a estos, ya que las comunidades de usuarios pueden promover mejoras en la educación promoviendo cambios en las políticas públicas de forma participativa. También la publicación de datos puede ser de utilidad para las instituciones educativas, porque aprender a analizar datos y desarrollar habilidades de análisis estadístico y de periodismo de datos puede ayudar a que los estudiantes participen activamente en los procesos democráticos. Para profundizar, les recomendamos ver este capítulo de la serie el Dato de Canal 13 Colombia sobre Datos Abiertos y Educación

**Energía y Medio Ambiente:** En este sector, que incluye a la academia, los datos sobre producción energética, contaminación, emergencias meteorológicas, desastres, y combustibles son imprescindibles, porque estos permiten mejorar el uso de las energías y reducir la contaminación, mediante cambios o mejoras en el transporte público, y para desarrollar formas eficientes de uso de energías.

**Corrupción y transparencia:** En relación con aquella información que es prioritaria para combatir la corrupción, como son los datos sobre finanzas, contratos, licitaciones, presupuesto y gasto en el sector público, estos le permiten a la sociedad civil, a la academia y a la ciudadanía saber de forma transparente dónde y cómo se gasta o invierte dinero del erario público. Estos datos son imprescindibles para poder monitorear la corrupción en los países, lo que incluye por ejemplo quién gana las licitaciones de productos y servicios, quienes son los proveedores del estado y quiénes y cómo manejan el dinero del estado.

**Crimen y justicia:** Los datos abiertos sobre crimen y justicia, pueden ser prioritarios para mejorar las políticas de seguridad ciudadana, pero también las políticas educativas y el uso de los recursos públicos, para mejorar el acceso a servicios y para dotar a las escuelas de más recursos. En el caso de estos datos, es importante considerar ciertos elementos éticos, porque su publicación, puede poner a ciertos grupos vulnerables en riesgo de exclusión y estigmatización, por lo que estos datos deben mostrar estadísticas que puedan apoyar la mejora de la calidad de vida de las personas y para asegurar el uso efectivo de los fondos públicos destinados a la policía.

# Introducción a los criterios para la apertura de datos

## Introducción

Es importante que las organizaciones conozcan los métodos para determinar la prioridad de la publicación de datos, por ejemplo, construyendo y co-creando un plan estratégico que contenga una serie de instrucciones, como una guía para la creación de políticas institucionales de datos abiertos. Estas sirven para que las organizaciones tengan procesos establecidos para definir qué conjuntos de datos priorizar en momentos determinados, fijando objetivos para los programas de datos abiertos, orientando la priorización en la publicación de conjuntos de datos, con calendarización de frecuencias, ya que esto es clave para los programas institucionales.

Estos objetivos deben cumplir con ciertas características. Por ejemplo:

1. Los datos de alto valor deben alcanzar al menos 3 estrellas en el esquema de 5 estrellas de Tim Berners-Lee
2. Valor para el propietario de los datos - Un conjunto de datos puede considerarse de alto valor cuando se cumplen uno o más de los siguientes criterios:
  - Datos que puedan contribuir a monitorear la transparencia ;
  - Datos que están sujetos a una obligación legal;
  - Datos que se relacionan directa o indirectamente con la función pública;
  - Datos que puedan ayudar a reducir costos o a mejorar operaciones

Tenemos que considerar que el valor de un conjunto de datos depende principalmente de su uso y potencial de reutilización, lo que puede conducir a la generación de actividad en torno a estos. El potencial del conjunto de datos está definido por:

- El tamaño y la dinámica de la audiencia objetivo.
- La cantidad de sistemas o servicios que podrían usar el conjunto de datos.
- Los conjuntos de datos que pueden tener un fuerte impacto social.

Es importante interactuar directamente con el público objetivo y las comunidades de usuarios para poder comprender el valor de su conjunto de datos, por lo que es necesario que las organizaciones que publican datos, establezcan canales de comunicación. Por ejemplo, informar a las comunidades de usuarios mediante una lista de correo, newsletters, mantener una página web actualizada con un canal de comunicación directa con las comunidades y tener canales activos en las redes sociales. También hay que considerar buscar medios de contacto y de comunicación efectiva con el público y las comunidades que no tienen acceso a internet, por ejemplo mediante foros con comunidades rurales, con adultos mayores o con usuarios que necesitan que las organizaciones les provean los datos en formatos accesibles, por ejemplo en braille, ya que es importante tener una comprensión compartida con la sociedad civil, la academia y los usuarios, de lo que se puede considerar como conjuntos

de datos de alto valor para que se pueda priorizar la publicación de estos conjuntos de datos.

## 2.1 Rol y niveles de consumo de los datos

El nivel de especialización de los usuarios puede determinar el valor de los datos, normalmente, los usuarios especializados analizan bases de datos, mientras que aquellos no-especializados por lo general consumen los datos a través de visualizaciones e historias. Lamentablemente, esta brecha de especialización en la comprensión de los datos, crea dos ejes o grupos: aquellos que tienen las capacidades de poder analizar y entender los datos en crudo, y que pueden generar análisis e historias y, quienes no tienen estas capacidades, conocimientos y herramientas, por ende, deben recibir los datos procesados, lo que los deja expuestos a ser manipulados o influenciados por aquellos grupos de poder que pueden contar historias con datos.

Una de las misiones que se deben contemplar en las estrategias de apertura de datos y acceso a la información pública, es incorporar programas para el desarrollo de capacidades en habilidades informacionales y de datos para reducir la brecha de alfabetización, para que los usuarios no-especializados puedan desarrollar capacidades para leer datos, entender la información presente en los datos y poder así desarrollar conclusiones a través del análisis de datos. Estos programas de capacitación deben incluir entrenamientos en el uso de tecnologías y de búsqueda de información, porque si no se promueve desde el sector público y desde el sector educacional el desarrollo de capacidades, la apertura de datos va a beneficiar solamente a quienes ya tienen las capacidades para aprovechar su potencial.

Otro elemento a considerar es la calidad de los datos, ya que el éxito de un portal de datos abiertos no está en el volumen de consultas de los datos, sino en el uso que se le da a estos por aquellos usuarios especializados que activamente juegan un rol de infomediarios. Por ejemplo, al transformar los datos a formatos y productos consumibles por usuarios no-especializados para resolver un problema o responder una duda, como también para la sociedad civil y la academia al realizar estudios e investigaciones.

Un ejemplo clásico de una plataforma exitosa de datos, que puede considerarse como infomediaria es el proyecto de <http://www.mejoratuescuela.org> que utiliza datos públicos sobre las escuelas de México para que padres y madres de familia puedan conocer información básica de las escuelas (nombre, dirección, contacto, etc.) pero también poder realizar análisis comparativos inter-escolares basados en los datos de los resultados de evaluaciones PISA y ENLACE de cada escuela. Este proyecto comprende que los padres y madres de familia no son usuarios especializados, no cuentan con tiempo ni capacidades para realizar descargas de datos públicos y posteriormente hacer el análisis para determinar qué escuela ha sido mejor calificada que otra en las evaluaciones de calidad educativa. Así, esta plataforma hace que cualquier usuario, de manera intuitiva, logre comparar escuelas sin necesidad de lidiar con la complejidad del acceso, limpieza y análisis de datos.

Comúnmente, las plataformas de datos de gobierno, sociedad civil y centros de investigación que buscan analizar temas específicos están dirigidas a públicos especializados, pues se asume que son quienes tienen interés particular en esos temas, que cuentan con el conocimiento y con las

capacidades necesarias para llegar a conclusiones sobre los datos, lo que puede marginar de la participación a aquellos que no poseen la expertise o capacidades necesarias para entender la información.

Un ejemplo de portales especializados, son los portales relacionados con presupuesto y gasto público en donde únicamente usuarios con conocimiento del tema pueden entender estos datos e información. Tal es el caso del portal de transparencia presupuestaria de México, <http://www.transparenciapresupuestaria.gob.mx/es/PTP/Presupuesto> de la Secretaría de Hacienda y Crédito Público de México o el sitio del paquete económico <http://paqueteeconomico.ciep.mx> del centro de investigación CIEP en México.

Para evitar que la brecha de acceso y comprensión a la información pública y a los datos se amplíe, es necesario que los gobiernos, sociedad civil y las universidades, hagan esfuerzos sustantivos y destinen presupuestos para que sus espacios informativos y de acceso a datos públicos puedan ser comprendidos por todas las personas mediante campañas de alfabetización y también sean intuitivos, accesibles, sencillos y consumibles por cualquier persona, teniendo expectativas realistas sobre el consumo de los datos.

Es crucial que los gobiernos, sociedad civil y las universidades, generen las herramientas y estrategias necesarias para formar a usuarios, para que todas las personas interesadas cuenten con el conocimiento y capacidades para poder participar de los procesos democráticos. Un ejemplo de este tipo de esfuerzos, es el que ha hecho Escuela de Datos en México, quienes han desarrollado un curso masivo abierto (MOOC) para desarrollar capacidades en temas de comprensión de la información sobre transparencia presupuestaria, en alianza con el gobierno y la sociedad civil, para permitir que cualquier persona pueda aprender sobre presupuestos <https://es.schoolofdata.org/2017/07/29/como-entender-el-presupuesto-caso-de-exito-de-un-mooc-en-mexico/>

## 2.2 Formatos por nivel de consumo de los datos

Si bien existen formatos que son más intuitivos para los distintos niveles de especialización de usuarios, así como distintos objetivos de comunicación, la tendencia en narrativas es poder combinar distintos elementos que puedan llevar a cada grupo de usuarios a consumir la información relevante.

Las agencias que proveen los datos, tienen el reto de diseñar espacios de información que sean útiles para usuarios especializados y no-especializados. Este reto obliga a establecer funcionalidades más precisas para ambos grupos de usuarios, entre las que hay que considerar las siguientes estrategias:

- Informar de forma clara a los usuarios especializados y hacer visible, mediante botones, menús o destacados, que existen espacios especializados, documentación de apoyo y zonas de descarga de datos.
- Informar a los usuarios no-especializados mediante información clara, imparcial y sencilla, para evitar abrumarlos si no pueden entender o acceder a los contenidos del sitio, o si notan demasiada complejidad al llegar a una sección, infografía o especialidad.
- Incluir estrategias para poder informar de forma clara, imparcial y sencilla a los usuarios que tienen necesidades especiales, incorporando sistemas de accesibilidad a la información, para prevenir la exclusión de

personas con discapacidades.

- Desarrollar o rediseñar el sitio o portal siguiendo las pautas básicas del diseño centrado en los usuarios (UX - user experience) para guiar a los usuarios por los temas, secciones y niveles de información en un espacio narrativo, ya sea impreso o digital;
- En caso de dudas sobre cuál es el consumo natural de la información, se recomienda iniciar de lo general a lo particular. Así son las conclusiones y definiciones básicas lo que primero se consume para que a medida que un usuario comprenda mejor el tema y los datos, pueda ir entrando y analizando el detalle hasta llegar a las secciones más especializadas como lo pueden ser los documentos de sustento original y secciones de descarga de datos abiertos.
- Intentar explicar los datos contextualizados en historias que sean de interés o que puedan ser relevantes para las personas a través de relatos utilizando técnicas de periodismo que incluyan no solo datos, sino también testimonios, o historias de la vida cotidiana.
- Proveer a los usuarios información de forma imparcial y sencilla, inclusive cuando los reportes tocan temas especializados o complejos, para evitar alienar a públicos no especializados, considerando que para este grupo es más difícil relacionarse con la información que para públicos especializados en datos, por que carecen de las habilidades necesarias para comprender información presentada de forma compleja.

Algunos ejemplos de historias basadas en datos, como las que presentamos a continuación, pueden servir para proyectar portales o sitios que presenten historias basadas en datos para poder entender ¿cuál es el rol de los datos y para qué nivel de especialización de audiencia va dirigida?

Estudios universitarios y salarios en Costa Rica:

<http://www.nacion.com/gnfactory/especiales/2015/carreras/desktop.html>

Pensiones en México:

<http://pensionesenmexico.ciep.mx/>

SOS Riachuelo:

<http://www.lanacion.com.ar/1937766-sos-riachuelo#Contaminaci%C3%B3n>

Obra Pública:

[http://transparenciapresupuestaria.gob.mx/es/PTP/Obra\\_Publica\\_Abierta](http://transparenciapresupuestaria.gob.mx/es/PTP/Obra_Publica_Abierta)

Medicamentos:

<http://medicamentalia.org/es/>

Desigualdad:

<https://calculadora-de-la-desigualdad.ojo-publico.com/>

## Introducción al desarrollo de planes de publicación de Datos Abiertos:

### Introducción

Para desarrollar un plan sustentable de publicación de datos, tenemos que enfocarnos en los datos prioritarios, en las capacidades organizacionales para abrir datos, las capacidades de nuestro público objetivo, en las metas que queremos alcanzar, y en el valor de los servicios que vamos a proveer. Por ende, hay una serie de elementos y estrategias a considerar que requieren de la participación de diversos actores.

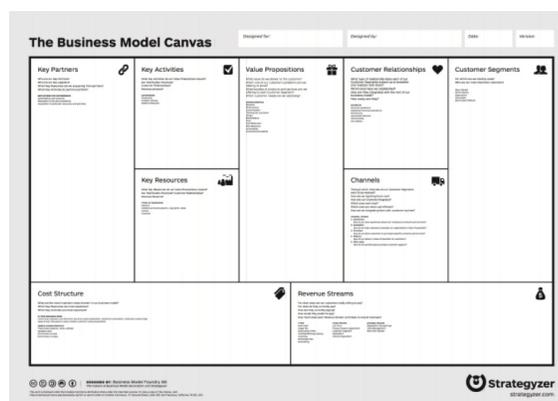
En esta unidad, les presentamos la herramienta Canvas, la cual debe ser incluida como parte del trabajo final, el cual se compone de una serie de elementos, las que incluyen un template de un plan de trabajo, una herramienta para el análisis FODA y un Canvas. En el caso del canvas, esta herramienta les va a permitir aproximarse al diseño de un plan de publicación de datos desde diversas perspectivas, considerando diversas interacciones e incluyendo a diversos actores.

### 3.1 Canvas: Elementos relevantes

El general, los canvas deben considerarse como una herramienta útil para desarrollar un proyecto o política de datos, y si bien hay variadas herramientas y mecanismos, en el contexto de este curso, nos hemos decantado por su uso, en conjunto con otras herramientas para promover la co-creación de planes de apertura.

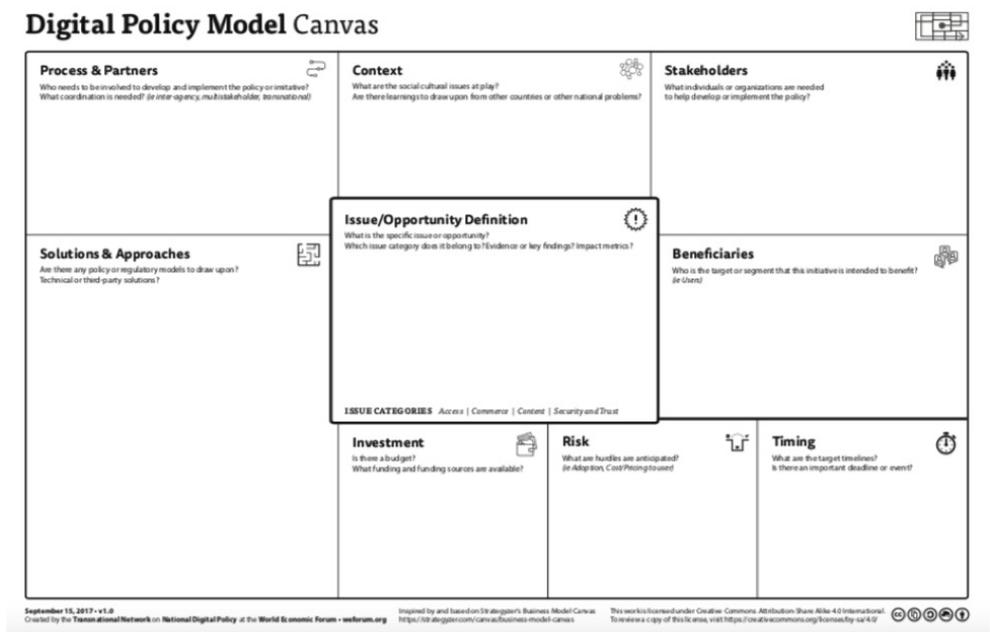
Al momento de formular tu proyecto institucional, te pediremos que uses la metodología CANVAS, la cual es frecuentemente utilizada en ámbitos de emprendimiento como herramienta para crear y diseñar modelos de negocio de una forma simple, sencilla, organizada y dinámica. Esta metodología permite ser adoptada en diferentes contextos, desde la creación de políticas institucionales, actividades didácticas, planes de trabajo colaborativos o proyectos de apertura de datos. El canvas original, es parte libro Business Model Generation de Alexander Osterwalder e Yves Pigneur, su objetivo es, que mediante es una plantilla de gestión estratégica, las organizaciones puedan desarrollar nuevos modelos de negocio o documentar los ya existentes. Es un gráfico visual con elementos que describen propuestas de producto o de valor de la empresa, la infraestructura, los clientes y las finanzas.

Fuente:  
Wikimedia  
Business Model Canvas



Otro modelo de Canvas es el Digital Policy Canvas (Políticas Digitales) desarrollado por el GovLab, el cual ha sido diseñado para desarrollar y apoyar la implementación de políticas digitales, considerando un problema específico y / o contexto geográfico, e incluye los siguientes pasos - elementos: primero, es necesario comenzar con un problema bien desarrollado o con la definición de una oportunidad, mayormente informada a través de consultas públicas, ya que es necesario, establecer un sentido claro para los beneficiarios, comprendiendo sus necesidades, aplicando metodologías de diseño centradas en el usuario.

Es importante para utilizar esta herramienta que se haga un mapa de potenciales participantes y se seleccione un grupo de partes interesadas clave, para involucrarlas en el desarrollo e implementación de la política, teniendo en consideración el contexto cultural, político y social de estas, adoptando un proceso inclusivo e iterativo, en el que hay que coordinar con diversos actores, evaluando y experimentando potenciales soluciones que involucren a diferentes personas para idear nuevos enfoques.



En general, la metodología establece una serie de bloques que, al completarse, visibilizan los elementos necesarios para poder desarrollar un proyecto exitoso, los cuales contienen los siguientes elementos y preguntas:

1. Segmentación de usuarios: ¿A quién le aporta el proyecto? ¿Qué segmentos engloban usuarios? ¿Cuáles son prioritarios? ¿Se busca contar con usuarios a nivel masivo, o bien diversificados o segmentados?
2. Propuesta de valor: ¿Qué valor ofrece el proyecto a los usuarios? ¿Qué problemas soluciona? ¿Qué necesidades satisface? ¿Qué beneficios aporta? ¿Qué características tiene el proyecto: novedad, operabilidad, personalización, diseño, eficiencia, reducción de riesgo, acceso, usabilidad, etc.?
3. Canales: ¿A través de qué canales/medios se debe conectar y atender a los usuarios? ¿Cómo están integrados los canales existentes? ¿Qué canales o medios funcionan mejor para nuestros usuarios / clientes? ¿Qué canales o medios son los más eficientes y eficaces para nosotros?

- ¿Qué canales o medios son más naturales, cercanos o intuitivos para nuestros clientes / usuarios?
4. Relación con usuarios: ¿Qué tipo de relaciones esperan los usuarios que se establezcan y se mantengan con ellos? ¿Qué relaciones están ya establecidas? ¿Cómo se integran en el modelo del proyecto? ¿Qué implica y cuánto cuesta iniciar y mantener las relaciones?
  5. Fuentes de financiamiento: ¿Cuánto cuesta implementar el proyecto? ¿Qué costo tiene el abrir los datos? ¿Cuál es el impacto de abrir datos que nadie va a usar?
  6. Recursos estratégicos: ¿Qué recursos son estratégicos para implementar y ejecutar el proyecto? ¿Los recursos requeridos son físicos, humanos, intelectuales, tecnológicos, financieros, etc.? ¿Qué recursos requiere el proyecto? ¿Cómo están vinculadas las fuentes de financiamiento con las relaciones requeridas?
  7. Actividades clave: ¿Qué actividades clave se pueden realizar? ¿Las actividades pueden ser: producción, solución de problemas, transaccionales, comunicación, involucramiento, formación, incidencia, etc.? ¿De qué manera se deben llevar a cabo?
  8. Alianzas y vínculos estratégicos: ¿Quiénes son nuestros socios naturales y cuáles aún deben serlo? ¿Quiénes deben ser los proveedores? ¿Con quién tendríamos que tener alianzas estratégicas? ¿Qué actividades deben realizar o apoyar cada contraparte? ¿Qué motiva, requiere o es de interés para cada contraparte?
  9. Estructura de Costos: ¿Cuáles son los costos inherentes al proyecto para su diseño, implementación y operación? ¿Qué recursos y actividades son las más costosas? ¿El proyecto está más basado en valor o costo? ¿Cuál es la naturaleza de los costos del proyecto: fijos, variables, escala, alcance, etc.?

### 3.2 Cómo utilizar el Canvas

Finalmente, para utilizar el Canvas para desarrollar proyectos, deben descargarse uno de los dos Canvas disponibles (uno para profesionales del servicio civil y otro para académicos y profesores universitarios). Para utilizar el Canvas, hay que asegurarse de poder responder a estas preguntas:

**1. Dimensiones de los usuarios** ¿Tienes un grupo de usuarios homogéneo o heterogéneo? Si tienes un grupo heterogéneo, ¿cuáles son sus segmentos?

**2. Composición de los segmentos:** Si las dimensiones de los segmentos son de tipo análisis 'macro', entonces se debe mirar dentro de cada grupo como 'Personas' (micro). Es necesario visualizar a estas Personas, ¿qué información requieren? ¿qué capacidades tienen? Y se debe comprender qué están buscando en sus portales y sitios Web.

**3. Problemas, necesidades, hábitos y alternativas actuales** ¿Qué trabajo estás haciendo por los usuarios? ¿Qué necesidades cubre tu proyecto? Asegúrate de identificar una necesidad / problema existente e identifica las alternativas específicas que el usuario utiliza actualmente. Si no está seguro, sal y observa, habla con personas representativas para poder vincular claramente el proyecto a las necesidades del público.

# Introducción al diseño centrado en las necesidades de los usuarios

## Introducción

Para que un proyecto de datos sea sustentable, y para que tenga un aporte real en las comunidades, es necesario que el diseño se centre en los usuarios y en sus necesidades, porque el costo de abrir datos que nadie va a usar, es muy alto para los estados. En esta unidad, el enfoque es pensar cuáles son los elementos que pueden beneficiar a los usuarios y a las organizaciones para que el impacto del proyecto sea real.

### 4.1 ¿Qué necesidades que tiene el usuario se buscan solucionar?

Todo proyecto de datos debe tener claro dos elementos: sus usuarios y sus objetivos. Muchas veces se comente el error de pensar que todas las personas tienen las mismas necesidades o que demandan cierta información. El asumir esas necesidades sin conocer en detalle a los usuarios o clientes a los que se quiere ayudar, generalmente hace que los esfuerzos de generación, uso y comunicación de datos sean poco o nada utilizados. Conocer las necesidades de los usuarios debe de ser un proceso continuo por lo menos en tres niveles:

**Informativos:** ¿qué información requiere un usuario en su vida cotidiana, proceso determinado o para resolver un problema? ¿en qué momento es que cada usuario requiere esa información? ¿qué medio es el más adecuado para cada usuario para recibir o consultar esa información?

**Servicios:** ¿qué servicios o transacciones son posibles de automatizar? ¿qué servicios pueden ser más cómodos, prácticos y eficientes para la vida, actividades y necesidades de cada usuario? ¿qué medio acerca y facilita más al usuario para realizar un servicio?

**Interacción y Respuesta:** ¿qué preguntas son las más comunes por parte de cada usuario? ¿en qué momento, actividad, circunstancia o proceso es que cada usuario requiere una respuesta o apoyo? ¿es posible automatizar las respuestas que requiere cada usuario? ¿cuándo se requiere que existan personas que atiendan directamente las peticiones de cada usuario? ¿a través de qué medio es mejor que cada usuario reciba la respuesta?

Si bien nunca se podrá cumplir con todas las necesidades de cada grupo de usuario, el tener claro quiénes son, qué requieren y qué medios son los más apropiados para brindarles información, servicios y respuestas permite tener elementos para priorizar los esfuerzos de desarrollo tecnológico y comunicación interna y externa.

### 4.2 ¿Qué información se requiere para solucionar esa necesidad?

Toda solución requiere información específica para ser brindada al usuario, analizada y después compartida o respondida al usuario. Y ahora ya debemos conocer las características precisas de esa información, tales como: ¿de dónde viene? ¿en qué formato está? ¿qué características tiene? ¿qué nivel de detalle debe de tener?

Es necesario desarrollar distintas soluciones tecnológicas y de comunicación, y lo fundamental que en todo desarrollo se requiere es que las partes, funcionales o técnicas, que estén involucradas tengan claridad sobre cómo es que debe funcionar la solución. Siempre es recomendable que en la fase de diseño se cuente con por lo menos dos documentos que definan el funcionamiento, operativo y técnico, de la solución.

**Funcionamiento operativo:** Se debe realizar siempre el diseño funcional u operativo primero pues es el que determina aquellos pasos que los usuarios deben recorrer al usar y administrar la solución. Así, se tiene total claridad de la ruta de uso y expectativas que cada usuario debe de tener así como las acciones que quienes administran y gestionan la información y tecnología deben hacer. Documentar estas expectativas hace muy fácil que en una fase de prototipado y pruebas se pueda evaluar si la solución está cumpliendo o no su objetivo. Algunas de las preguntas más simples que se deben de responder en un documento de funcionalidad operativa son:

- ¿cuáles son los usuarios y qué deben hacer?
- ¿cuáles son las rutas o pasos que idealmente debe seguir un usuario?
- ¿qué elementos, guías y partes de la funcionalidad debe siempre cada usuario identificar y saber utilizar?
- ¿cómo se debe de asistir al usuario desde la administración técnica y de información?
- ¿qué debe de suceder en relación con el uso de la solución?

**Funcionamiento técnico:** Aunque en un escenario óptimo se cuenta con contrapartes técnicas para el desarrollo de una solución, es básico contemplar que una solución tecnológica implica tanto la construcción como adecuación de código y programas computacionales. Esto implica que, inclusive en el escenario de desarrollo más simple (como puede ser realizar un gráfico animado o interactivo), así como en lo más complejo (como en el caso de plataformas o sistemas de información) es necesario tener un documento que indique el flujo de la información y documentación del código, programas y complementos utilizados.

Esta documentación sirve para que, antes de iniciar los esfuerzos de desarrollo, haya total empate entre lo que se construirá con tal de cumplir con las funcionalidades establecidas en el documento operativo, y posteriormente sirva como documentación base para gestionar actualizaciones, mantenimientos y posteriores versiones de la solución. Si bien existen diversas metodologías para diseñar, documentar y desarrollar soluciones basadas en tecnología digital, es fundamental siempre tener claras las respuestas a las siguientes preguntas:

- ¿cómo funciona la tecnología elegida?
- ¿quién es responsable de la gestión, mantenimiento y actualización técnica y operativa?
- ¿cómo es que los datos de insumo se cargan en la plataforma? ¿quién es responsable de la carga y actualización de datos?
- ¿qué plataformas, lenguajes de programación, protocolos de ingreso de información y programas son la base de la solución?
- ¿qué mantenimiento de código, actualizaciones, seguridad, limpieza de datos, plug-ins o complementos requiere la solución?
- ¿en qué medida se puede escalar o aumentar las funcionalidades de

la plataforma en caso de que se requiera incrementar, diversificar o complementar su uso?

La siguiente lectura analiza, desde un híbrido de perspectivas técnicas y de usuario 10 elementos básicos que los sitios de gobierno requieren tomar en cuenta: <https://socialtic.org/blog/10-exigencias-basocas-para-sitios-de-gobierno/>. Igualmente, existen comunidades de práctica especializadas que nos permiten identificar y dar seguimiento a las características esenciales de distintos tipos de soluciones.

### 4.3 ¿Qué factores de éxito se deben considerar?

El éxito de una solución requiere siempre de factores ajenos a los datos y la tecnología. Muchas veces se pueden identificar soluciones que en teoría aparentan ser óptimas pero que cuando se analiza su entorno institucional, social y técnico, hacen que el proyecto sea mucho más complejo o inclusive inviable.

Antes de invertir y desarrollar soluciones basadas en datos es fundamental hacer un análisis de factores externos con tal de identificar potenciales riesgos, acciones complementarias y esfuerzos indispensables que harán que la solución sea utilizada con éxito por sus usuarios. A continuación se identifican 5 rubros a analizar:

1. Políticas institucionales y normatividad apropiada
2. Acciones para el involucramiento ciudadano
3. Infraestructura tecnológica y/o física apropiada
4. Acceso, disponibilidad e interoperabilidad de la información
5. Coordinación y respuesta institucional a la ciudadanía

Para profundizar sobre el rol de los datos para la ciudadanía les recomendamos el siguiente video: <https://www.youtube.com/watch?v=lqxttLRYK4k>

Por ejemplo, desde 2008 se han implementado docenas de soluciones que potencialmente le permiten a la ciudadanía reportar o denunciar, desde servicios públicos en mal funcionamiento hasta hechos de violencia o riesgo. En retrospectiva, la experiencia derivada de estos proyectos ha hecho visible que su fracaso se debe principalmente a fallas en factores ajenos a los datos y la tecnología. Aquí se reflexiona sobre 4 factores esenciales para este tipo de soluciones: <https://socialtic.org/blog/4-respuestas-para-entender-y-abordar-proyectos-de/>

Es importante que al leer estas reflexiones, así como cuando diseñamos, construimos o utilizamos alguna solución, tengamos siempre en mente si se contemplaron adecuadamente los factores de éxito aquí identificados.

# Formulación de proyectos de datos abiertos a nivel organizacional

## Introducción

Para poder desarrollar un plan estratégico de apertura de datos, es necesario desarrollar un mapa de ruta para determinar qué elementos, recursos, participantes, guías, estructuras, consultas y organización necesitan para desarrollar un proyecto sustentable que ayude a resolver problemas relacionados con las necesidades de los usuarios.

### 5.1 Promover la demanda y la participación

Durante la etapa inicial, las estrategias de participación se deben centrar en identificar a los principales proveedores y consumidores de datos, y alentarlos a participar en co-crear una estrategia de apertura en una iniciativa de datos abiertos. En el caso de los proveedores, la participación tiene que consistir en identificar a los administradores de datos de los principales organismos del Gobierno, ya que a veces, los departamentos del sector público, necesitan ayuda para entender qué datos ya tienen en su poder, cuáles de estos son prioritarios, cuáles están abiertos pero en formatos inaccesibles y necesitan ser reformateados, qué información está disponible en otros departamentos, y qué información está disgregada entre varias unidades.

Para poder entender a qué necesitan acceder y entender los diversos grupos de usuarios, las estrategias para fomentar participación deben ser diseñadas para identificar instrumentos que sean útiles para que las organizaciones, la sociedad civil, las agrupaciones sociales, la academia y las personas, desarrollen capacidades mediante incentivos y que cuenten con los medios para usar los datos abiertos.

En este contexto, la participación temprana por lo general implica mayor conciencia sobre el uso, los beneficios y las oportunidades que presentan los datos abiertos, la identificación de las prioridades y necesidades de acceso a datos para distintos públicos, el fortalecimiento de la capacidad, y el respaldo a las comunidades de datos abiertos que ya existen. Los siguientes eventos y programas son ejemplos de estrategias eficaces:

#### **Talleres de concienciación sobre datos abiertos.**

Estos eventos, diseñados para una variedad de audiencias, ofrecen un panorama general de las mejores prácticas y los beneficios de los datos abiertos, y pueden incluir aportes de otros países, regiones y organizaciones sobre sus propias experiencias.

#### **Mesas redondas sobre el valor de los datos abiertos.**

Las mesas redondas ofrecen oportunidades para presentar al público los beneficios de los datos abiertos para el crecimiento y el desarrollo social, y están orientadas para aprovechar diferentes liderazgos, tanto en el lado de la oferta como en el de la demanda. La información proporcionada en esta mesa redonda se debe adaptar a las necesidades de los grupos encargados de la toma de decisiones—incluidas las autoridades ministeriales, los líderes de los Gobiernos locales, los directores de organizaciones de la sociedad civil, ciudadanos y académicos y otros grupos de usuarios.

### **Maratones para liberar datos**

Es un programa que reúne a codificadores para recolectar datos gubernamentales existentes pero no estructurados, reestructurarlos en formatos que se puedan leer con máquinas (por ejemplo, CSV) con el fin de alimentar una plataforma de datos. Es importante considerar que estas maratones no pueden basarse únicamente en trabajo voluntario, por ejemplo de estudiantes, por lo que hay que buscar mecanismos para proveer a los participantes con retribuciones y compensaciones por el trabajo realizado.

## **1.2 Elementos a Considerar**

Para que los datos sean útiles, los usuarios primero deben adquirir conocimientos sobre estos: deben entenderlos, saber cómo obtenerlos y cómo aplicarlos para resolver problemas o solucionar necesidades específicas. Las estrategias de participación deben apuntar a aumentar los conocimientos sobre datos y la capacidad técnica de los posibles usuarios, incluidos periodistas, organizaciones sin fines de lucro, sociedad civil, sector privado, académicos y desarrolladores.

El grado y estilo de capacitación técnica que deben tener los programas de capacitación deben estar alineados con el nivel de las audiencias. Idealmente, los usuarios deberían tener acceso a múltiples oportunidades de capacitación, de acuerdo con sus necesidades, y conectarse con grupos de interés y comunidades de práctica locales.

Algunos ejemplos de métodos, cursos y talleres que abordan estos objetivos son, con diversos niveles de profundidad y de formas se pueden ver a continuación:

### **Centros de entrenamiento sobre datos abiertos**

Estos talleres intensivos de aprendizaje práctico se centran en el desarrollo de aptitudes básicas para periodistas, la sociedad civil, la academia y los codificadores cívicos. Cada participante debe tener acceso a un computador y a sets de datos para participar en trabajos prácticos guiados. En las sesiones se deben incluir demostraciones de catálogos de libre acceso; capacitación sobre Excel/planillas de cálculo y la creación de cuadros dinámicos; extracción de datos a nivel local y datos nacionales; filtrado de estos datos; creación de visualizaciones; una introducción a mapas y datos de geolocalización, y redacción de textos basados en los datos; también puede incluir un programa de introducción a la creación de aplicaciones y planes de acción.

### **Clases magistrales sobre datos abiertos**

Estas clases están destinadas a especialistas del lado de la demanda que ya manejan el tema y que aspiran a adentrarse aún más en el aprendizaje práctico y la formación de aptitudes con análisis de datos.

### **Cursos online sobre datos abiertos**

Estos curso sobre datos abiertos, sirven para proveer a diferentes comunidades de usuarios un conjunto amplio y contextualizado de herramientas sobre manejo de datos.

### 5.3 Utilización y reutilización

El objetivo de esta etapa es aprovechar los datos abiertos de manera específica para abordar desafíos y oportunidades particulares, y utilizarlos para generar valor y resultados tangibles. Es posible que los Gobiernos ya tengan prioridades de políticas para las que buscan soluciones relacionadas con los datos abiertos, y la participación en esta etapa puede estar diseñada para empezar a formular esas soluciones. En otros casos, el sector privado o la academia, pueden tener sus propias ideas y el objetivo puede ser simplemente fomentar la innovación hacia cualquier propósito público. Los dos tipos de eventos que se enlistan a continuación, son ejemplos de actividades que se orientan a desafíos específicos y tienen el objetivo de encontrar las soluciones apropiadas a diversos problemas:

**Hackatones:** Las hackatones se organizan en torno a desafíos de desarrollo específicos de los que ya se ha obtenido información. Habitualmente, los tecnólogos usan los datos existentes para generar y crear aplicaciones, que resuelvan problemas de desarrollo, haciendo hincapié en la prestación de servicios públicos. Los ganadores de una hackatón pueden recibir retribuciones económicas, o bien capitales iniciales, para crear y poner en funcionamiento aplicaciones.

**Fondos prototipo:** Un proceso de alianza para conectar proyectos bien formulados sobre aplicaciones de servicios y uso de datos con financiamiento para construir y aplicar modelos ampliables.

### 5.4 Desarrollo de un plan de publicación de Datos Abiertos

Para partir, veamos un ejemplo:

#### **Desarrollar un sistema centralizado de Contrataciones Públicas** (3er PAN 2016-2018-Bulgaria)

**¿Cuál es el Compromiso?:** Desarrollar un sistema centralizado de contratación pública que contenga todas las fases del proceso y que habilite la presentación electrónica de ofertas.

**Problemática:** “La contratación pública es un tema delicado para la sociedad búlgara. Actualmente el proceso de contratación pública no está informatizado ni estandarizado, lo que hace necesario que los interesados preparen la documentación de licitación en papel y la presenten en las oficinas de la autoridad contratante, lo cual es ineficaz, lento y conlleva un costo significativo mientras al mismo tiempo, dificulta el procesamiento y la evaluación de las ofertas para la administración”

**¿Cómo contribuye a resolver la problemática?:** Simplificará los procedimientos de contratación pública y creará garantías para una evaluación transparente e imparcial de las ofertas y de las contrataciones.

#### **Hitos**

- Implementar un estándar claro de contrataciones.
- Apoyar a los adjudicadores proveyéndoles las herramientas- sistemas y plantillas.
- Publicar información en todas las etapas de la adquisición.
- Mantener actualizado el registro público de contratos gubernamentales.

Para poder desarrollar un plan de publicación de datos abiertos, hay que determinar un objetivo claro, y un público objetivo, y comenzar identificando aquellos conjuntos de datos que son aptos y necesarios para su publicación, considerando los requisitos de las partes interesadas a nivel interno y externo. El plan de publicación de datos abiertos debe desarrollarse teniendo en cuenta las necesidades de los interesados pertinentes, así como los posibles beneficios, riesgos y costos de la publicación de los conjuntos de datos.

Hay que considerar que los organismos del sector público tienen una gran cantidad de diversos conjuntos de datos. Sin embargo, también operan bajo restricciones presupuestarias y varias otras que les impiden hacer todos los conjuntos de datos posibles para su reutilización a la vez, por lo que cuando se desarrolle un plan de publicación de datos abiertos, es importante equilibrar los requisitos y las restricciones conocidas. El plan de publicación de datos abiertos debe contener todas las tareas necesarias para implementar la iniciativa de datos abiertos y debe ser respaldado por una hoja de ruta para la publicación de los conjuntos de datos seleccionados.

Con respecto a la hoja de ruta de publicación, la publicación de los conjuntos de datos debe priorizarse teniendo en cuenta los requisitos de los interesados, así como los resultados del análisis de los conjuntos de datos, es decir, los beneficios y riesgos identificados y las estimaciones de costo / esfuerzo. El progreso debe rastrearse según el plan de publicación de datos abiertos, el cual debe revisarse y actualizarse regularmente.

Es importante enfocar la publicación de datos abiertos, ya que esta debe estar en línea con los objetivos estratégicos de la organización, con las políticas relevantes, como la estrategia nacional de Datos Abiertos, y debe satisfacer la demanda de los usuarios potenciales de los datos. Por lo tanto, se deben identificar y analizar los requisitos de las partes interesadas internas y externas. El plan de publicación de datos abiertos debe contener todas las tareas necesarias para implementar la iniciativa de datos y los pasos de implementación se pueden resumir de la siguiente manera:

- Obtener el apoyo de la dirección de la unidad, ya que el compromiso de las jefaturas para otorgar recursos y disponer de espacios es clave para poder desarrollar un plan de publicación de datos abiertos sustentables.
- Designar a un equipo multidepartamental responsable del desarrollo del plan de publicación de Datos Abiertos.
- Crear mecanismos de cooperación y líneas de comunicación estratégica para informar a los actores internos y a los departamentos que deben ser involucrados del proyecto de datos abiertos.
- Diseñar líneas y canales de comunicación estratégica para informar a los actores externos de los progresos y de las etapas de trabajo del proyecto de datos abiertos, y para identificar los requisitos y necesidades de los interesados externos.

## MÓDULO 5

# Lecturas

Datos Abiertos: Guía estratégica para su puesta en marcha Conjuntos de datos mínimos a publicar <http://femp.femp.es/files/3580-1617-fichero/Gu%C3%ADa%20Datos%20Abiertos.pdf>

Guía rápida para apertura de Datos Gubernamentales <http://sgp.pcm.gob.pe/wp-content/uploads/2016/10/Guia-Rapida-ADG.pdf>

Atenas, J.. (2016, October 5). Desarrollo de un modelo docente en el uso Datos Abiertos como Recursos Educativos Abiertos para América Latina: La experiencia de la Universidad de la República, Uruguay (Version 1). figshare. <https://doi.org/10.6084/m9.figshare.3984508.v1>

Atenas, J, & Havemann, L. (2018). Open Data & Education (Version pre-print). In State of Open Data. African Minds. <http://doi.org/10.5281/zenodo.1544942>

El ecosistema de Gobierno Abierto en la Ciudad de Buenos Aires: Un modelo generador de cambios. Castagnola, Andrea; Herrero, Álvaro: <https://publications.iadb.org/es/el-ecosistema-de-gobierno-abierto-en-la-ciudad-de-buenos-aires-un-modelo-generador-de-cambios>

## Agradecimientos



## Sobre la editora

**Javiera  
Atenas**

Doctora en Educación y senior teaching fellow de la Higher Education Academy del Reino Unido. Investigadora Principal en la Iniciativa Latinoamericana de Datos Abiertos (ILDA).