



Manual de los Datos Abiertos



Open Knowledge Foundation

Un proyecto de la Fundación para el Conocimiento Abierto.

©2010-2012, Open Knowledge Foundation.

Adaptado, corregido y ampliado por *Carlos Brys*

Esta obra está bajo una Licencia

Creative Commons Atribución v3.0.



Licencia

Este manual es una obra derivada de los documentos:

Open Data Handbook Documentation y

El manual de Open Data

de la **Open Knowledge Foundation**

Los documentos originales pueden obtenerse en la dirección de Internet:

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

que se distribuyen bajo una licencia

Creative Commons Atribución v3.0.

Adaptado, corregido y ampliado por *Carlos Brys*

carlos.brys@gmail.com

@CarlosBrys

Créditos

Autores

- Daniel Dietrich
- Jonathan Gray
- Tim McNamara
- Antti Poikola
- Rufus Pollock
- Julian Tait
- Ton Zijlstra

Fuentes utilizadas

- Propuesta técnica para la implementación de la Iniciativa Internacional para la Transparencia de la Ayuda (IATI, International Aid Transparency Initiative). Grupo Consultivo Técnico de IATI dirigido por Simon Parrish
- Abriendo el potencial de la información de ayuda. Rufus Pollock, Jonathan Gray, Simon Parrish, Jordan Hatcher
- Manual finlandés escrito por Antti Poikola
- Reporte Más allá del acceso. Información de acceso y Fundación para el Conocimiento Abierto

Otras fuentes

W3C Publishing Government Data (2009)

<http://www.w3.org/TR/gov-data/>

El Manual de los Datos Abiertos

Este manual se refiere a los aspectos legales, sociales y técnicos de la apertura de datos. Puede ser utilizado por todos pero fue específicamente diseñado para aquellos que están en la búsqueda de la apertura de datos.

Referencia el *por qué*, el *qué* y el *cómo* del **Open Data**. Por qué abrir, qué es abrir y cómo abrir los datos.

Damos una cálida bienvenida a los comentarios sobre el texto e incorporaremos retroalimentación conforme avancemos.

También damos la bienvenida a contribuciones y sugerencias para secciones adicionales y áreas a examinar.

Si usted está interesado contribuir con material, por favor ponerse en contacto a esta dirección:

opendatahandbook@okfn.org

Índice general

1. Introducción	1
2. ¿Por qué Datos Abiertos?	5
3. ¿Qué son los Datos Abiertos?	11
4. ¿Cómo Abrir los Datos?	15
5. Ya he Publicado Cierta Información, ¿y ahora qué?	31
A. Glosario	43
B. Formatos de Archivos	49
C. ¿Qué derechos legales (IP) hay en las Bases de datos?	59

Introducción

¿Tu sabes cuanto del dinero de tus impuestos es destinado en iluminación de las calles o en estudios sobre el cáncer?. ¿Cuál es el camino más corto, seguro y pintoresco para ir en bicicleta de tu casa al trabajo?. ¿Donde puedes encontrar la mejor oportunidad laboral y el número más alto de árboles frutales per cápita?. ¿Cuándo puedes influir en decisiones sobre tópicos que te preocupan profundamente, y con quien debes hablarlo?

Ahora, las nuevas tecnologías hacen posible crear servicios para responder estas preguntas automáticamente. Mucha de la información necesaria para responder estas preguntas es generada por organismos públicos. Sin embargo, a menudo los datos requeridos aún no están disponibles en una forma fácil de usar. Este libro es acerca de cómo liberar el potencial de la información oficial y de otro tipo para permitir nuevos servicios, para mejorar las vidas de los ciudadanos y para hacer el trabajo del gobierno y la sociedad mejor.

La noción del término “Datos Abiertos” y específicamente de “Datos Abiertos del Gobierno” - información, publica o no, que cual-

quiera es libre de acceder y reutilizar para cualquier propósito - ha estado circulando por varios años. En 2009 los datos abiertos comenzaron a hacerse visibles en la corriente principal, a través de varios gobiernos (como los EE.UU., Reino Unido, Canadá y Nueva Zelanda) que anunciaron sus iniciativas hacia la apertura de su información pública.

Este libro explica los conceptos básicos de “Datos Abiertos”, especialmente en relación con el gobierno. Cubre cómo a partir de datos abiertos se crea valor y se puede tener un impacto positivo en muchas áreas diferentes. Además de explorar los antecedentes, el manual también provee información concreta sobre cómo producir datos abiertos.

Dirigido a

Este manual tiene una amplia audiencia:

- Para ellos que jamás han escuchado anteriormente sobre datos abiertos y aquellos que se consideran experimentados profesionales de los datos
- funcionarios y activistas
- para periodistas e investigadores
- políticos y desarrolladores
- geeks de los datos y los que nunca han oído hablar de una API.

La mayoría de la información ofrecida actualmente esta centrada en datos del sector público. Sin embargo, las intenciones de los au-

tores son más amplias si el tiempo lo permite. Eres bienvenido para participar y ayudarnos en este esfuerzo.

El manual apunta a aquellos con poco o sin conocimiento sobre la temática. Si te encuentras con jerga o terminología con la que no estás familiarizado, por favor recurre al Glosario y a las PF (Preguntas Frecuentes) que puedes encontrar al final del manual.

Capítulo 2

¿Por qué Datos Abiertos?

Los *Datos Abiertos*, especialmente los *Datos Abiertos del Gobierno*, son un gran recurso que todavía en gran parte están sin explotar. Muchas personas y organizaciones recogen una amplia gama de diferentes tipos de datos con el fin de utilizarlos sus tareas. El gobierno es particularmente significativo en este sentido, tanto por la cantidad y la centralidad de los datos que recoge, sino también porque la mayoría de los datos del gobierno por ley son públicos, y por lo tanto podrían ser abiertos y estar disponibles para que otros los puedan usar.

¿Por qué esto es de interés?

Existen muchas áreas donde podemos esperar que los datos abiertos sean valiosos, y donde los ejemplos de cómo ha sido usado ya existan. Existen también diferentes grupos de personas y organizaciones que pueden beneficiarse de la disponibilidad de datos abiertos, incluido el mismo gobierno. Al mismo tiempo es imposible predecir con exactitud cómo y donde será creado el valor en el futuro.

La naturaleza de las innovaciones es que su desarrollo frecuentemente proviene de los lugares más inverosímiles.

Ya es posible apuntar a un gran número de áreas donde los gobiernos abiertos están creando valor. Algunas de estas áreas incluyen:

- Transparencia y control democrático
- Participación
- Auto-empoderamiento
- Productos privados y servicios nuevos o mejorados
- Innovación
- Mayor eficiencia de los servicios públicos
- Mayor eficacia de los servicios públicos
- Medición del impacto de políticas
- Nuevos conocimientos a partir de fuentes de datos combinadas y patrones en grandes volúmenes de datos

Existen ejemplos en la mayoría de estas áreas.

En términos de transparencia, proyectos como el finlandés “*tax tree*” y el británico “*Where does my money goes?*” muestran cómo el Gobierno está gastando el dinero de tus impuestos. También está el ejemplo de cómo los Datos Abiertos hicieron que el Gobierno de Canadá ahorre 3.2 millones de dólares que de manera fraudulenta se deducían de los impuestos como donaciones de caridad. Varios sitios web, como el danés *folketsting.dk*, siguen las actividades en el

Parlamento y el proceso de formulación de leyes para que puedas ver exactamente qué está pasando y que Miembros del Parlamento estén involucrados.

Los datos abiertos también pueden ayudarte a tomar mejores decisiones en tu vida, o a tener un rol más activo en la sociedad. Una mujer en Dinamarca desarrolló *findtoilet.dk* que muestra todos los baños públicos daneses, para que aquellas personas que sabían tenían problemas de vejiga pudieran salir más seguido. En los Países Bajos está disponible un servicio, *vervuilingsalarm.nl*, que te avisa con un mensaje si la calidad del aire en tus inmediaciones alcanzará, al día siguiente, el umbral que previamente definiste. En Nueva York puedes saber fácilmente dónde puedes pasear a tu perro, así como también encontrar otras personas que usan los mismos parques. Servicios como “*Mapumental*” en el Reino Unido y “*Mapnificent*” en Alemania te permiten encontrar lugares donde vivir tomando en cuenta la duración del viaje hasta tu trabajo, precios de viviendas y cuán bella es el área. Todos estos ejemplos usan Datos Abiertos.

En términos económicos, los datos abiertos son también de gran importancia. Numerosos estudios estimaron el valor económico de los datos abiertos en varias decenas de billones de Euros al año, sólo en la Unión Europea. Nuevos productos y compañías están reutilizando datos públicos. El sitio danés *husetsweb.dk* te ayuda a encontrar maneras de mejorar la eficiencia energética en tu casa, incluyendo planificación financiera e información sobre constructores que puedan hacer el trabajo. Está basado en información catastral y sobre subsidios gubernamentales, así como el registro de comercio local. El Traductor de Google usa el enorme volumen de documen-

tos de la Unión Europea que aparecen en todos los idiomas europeos para entrenar sus algoritmos de traducción y así mejorar la calidad de su servicio.

Los datos abiertos son de valor también para el mismo Gobierno. Puede mejorar, por ejemplo, la eficiencia gubernamental. El Ministro de Educación holandés publicó en internet todos los datos relacionados con educación, para su reutilización. Desde entonces, el número de preguntas que reciben bajó, reduciendo el volumen de trabajo y los costos, y las preguntas restantes resultan más fáciles de responder porque es claro dónde puede encontrarse la información importante. Abrir datos es también hacer al Gobierno más efectivo lo que, en última instancia, reduce costos. El Departamento Holandés de Patrimonio Cultural esta liberando información y colaborando con sociedades y grupos históricos amateurs, como la Fundación Wikimedia, con el objeto de ejecutar sus propias tareas de manera más eficiente. Esto no solamente deviene en mejoras en la calidad de sus datos, sino que también terminará haciendo a los departamentos más pequeños.

Mientras que existen diversas instancias que demuestran cómo los datos abiertos están creando valor tanto social como económico, todavía no sabemos qué cosas se volverán posibles en el futuro. Nuevas combinaciones de datos pueden crear nuevos conocimientos e ideas, que pueden llevar a nuevos campos de aplicación. Hemos visto esto en el pasado, por ejemplo cuando el Dr. Snow descubrió la relación entre tomar contaminación de agua y contraer cólera en Londres en el siglo XIX, combinando información de muertes por cólera y la ubicación de los pozos de agua. Esto llevó a la construcción del sistema de acantarillado de Londres y mejoró enorme-

mente la salud de la población. Es probable que veamos desarrollos como tal sucediendo de nuevo mientras ideas inesperadas surgen de la combinación de diferentes conjuntos de datos abiertos.

Este potencial sin explotar puede ser aprovechado si transformamos los datos públicos de Gobierno en datos abiertos. Esto solamente pasará, sin embargo, si se trata de datos verdaderamente abiertos, es decir si no hay restricciones (legales, financieras o tecnológicas) para su reutilización por terceros. Toda restricción excluirá a personas de la posibilidad de reutilizar los datos públicos y hará más difícil que se encuentren maneras valiosas de hacerlo. Para que el potencial sea reconocido, los datos públicos deben ser abiertos.

Capítulo 3

¿Qué son los Datos Abiertos?

Este manual es sobre “Datos Abiertos”, pero ¿qué son exactamente los Datos Abiertos?. En particular ¿qué hace a los Datos Abiertos sean “Abiertos”? ¿De qué tipo de datos hablamos?

¿Qué Significa Abierto?

Para nuestros propósitos, Datos Abiertos está definido por la “Definición de Apertura”

“Los datos abiertos son datos que pueden ser utilizados, reutilizados y redistribuidos libremente por cualquier persona, y que se encuentran sujetos, cuando más, al requerimiento de atribución y de compartirse de la misma manera en que aparecen”.

La definición de apertura completa da detalles precisos de lo que significa. Para resumir lo más importante:

- Disponibilidad y acceso: la información debe estar disponible como un todo y a un costo razonable de reproducción, prefe-

riblemente descargándola de internet. Además, la información debe estar disponible en una forma conveniente y modificable.

- **Reutilización y redistribución:** los datos deben ser provistos bajo términos que permitan reutilizarlos y redistribuirlos, e incluso integrarlos con otros conjuntos de datos.
- **Participación universal:** todos deben poder utilizar, reutilizar y redistribuir la información. No debe haber discriminación alguna en términos de esfuerzo, personas o grupos. Restricciones “no comerciales” que prevendrían el uso comercial de los datos; o restricciones de uso para ciertos propósitos (por ejemplo sólo para educación) no son permitidos.

Si te estás preguntando por qué es tan importante ser claros respecto de lo que significa ser “Abierto” y por qué se usa esta definición, hay una respuesta simple: *interoperabilidad*.

La interoperabilidad denota la habilidad de diversos sistemas y organizaciones para trabajar juntos (interoperar). En este caso, es la habilidad para interoperar o integrar diferentes bases de datos.

La interoperabilidad es importante porque permite que distintos componentes trabajen juntos. Esta habilidad de integrar componentes es esencial para construir sistemas complejos y grandes. Sin interoperabilidad es casi imposible lograrlo -como lo demuestra el famoso mito de la Torre de Babel, donde la inhabilidad de comunicar (interoperar) da como resultado el colapso de todo el esfuerzo para construir la torre.

Enfrentamos una situación similar en relación con los datos. La esencia de los datos (o códigos) compartidos es que una parte del

material abierto pueda a partir de ahí ser mezclado con otro material abierto. Esta interoperabilidad es absolutamente fundamental para entender los principales beneficios prácticos de la apertura: el incremento dramático de la habilidad de combinar distintas bases de datos o conjuntos de datos y así desarrollar más y mejores productos y servicios (estos beneficios son discutidos con más detalle en la sección del porqué de los datos abiertos).

Proveer una definición clara del concepto de apertura garantiza que cuando se trabaje con conjuntos de datos abiertos de diferentes fuentes, se los pueda combinar; y asegura que evitemos nuestra propia “Torre de Babel”: muchos conjuntos de datos con poca o ninguna posibilidad de combinarlos en sistemas más grandes, donde se encuentra el verdadero valor.

¿De qué Datos Estamos Hablando?

Los lectores ya vieron ejemplos de los tipos de datos que son o pueden ser abiertos y verán más ejemplos a continuación. Sin embargo, será de utilidad bosquejar que tipos de datos son o pueden ser abiertos e -igualmente importante- cuáles no serán abiertos.

La clave es que cuando se trata de abrir datos el foco se pone en información no personal, es decir, datos que no contienen información sobre individuos específicos.

De manera similar, para algunos tipos de datos gubernamentales, pueden aplicar restricciones nacionales de seguridad.

Capítulo 4

¿Cómo Abrir los Datos?

Esta sección constituye el núcleo de este manual. Le da consejos concretos y detallados de cómo los poseedores de datos pueden abrir sus datos. Pasaremos por lo más básico, pero también cubriremos las trampas. Por último, vamos a discutir las cuestiones más sutiles que pueden surgir.

Existen tres reglas clave que recomendamos seguir en la apertura de datos:

- **Mantenlo simple.**

Inicia con algo pequeño, sencillo y rápido. No hay requisitos que los conjuntos de datos deban tener en este momento. Inicia abriendo sólo un conjunto de datos (*dataset*), o incluso una parte de un conjunto grande de datos, estará muy bien, y claro, entre más conjuntos de datos puedas abrir, mejor.

Recuerda que esto es acerca de la innovación. Moverse tan rápido como puedes es bueno porque significa que puedes crear un impulso y aprender de la experiencia. La innovación es tanto

sobre el fracaso como el éxito y entiende que no todo dataset será útil.

- **Comprometerse temprano y comprometerse con frecuencia**

Participar con los usuarios reales y potenciales y reutilizadores de los datos tan pronto y tan a menudo como se pueda, ya sean ciudadanos, empresas o desarrolladores. Esto asegurará que la próxima versión de su servicio sea tan relevante como pueda ser.

Es esencial tener en cuenta que gran parte de los datos no llegarán a los usuarios finales directamente, sino a través de “infomediarios”. Estas son las personas que toman los datos y los transforman para presentarlos. Por ejemplo, la mayoría de nosotros no quiere o necesita una gran base de datos de las coordenadas GPS, preferiríamos un mapa. Por lo tanto, en primer lugar comprométase con los infomediarios. Ellos volverán a utilizar y reutilizar el material.

- **Direcciona temores y malentendidos comunes.**

Esto es especialmente importante si usted está trabajando con o dentro de las grandes instituciones como el gobierno. Cuando abra los datos se encontrará con un montón de preguntas y temores. Es importante (a) identificar las más importantes y (b) abordar estos problemas en la fase más temprana posible.

Existen cuatro pasos principales para la creación de Datos Abiertos, cada uno será explicado en detalle más abajo. Se encuentran en un orden muy aproximado y pueden ser realizados simultáneamente.

1. **Elegir el conjunto de datos que desea hacer público.**

Elija el conjunto de datos que desea hacer público. Tenga en cuenta que usted puede (y puede ser necesario) regresar a este paso si tiene problemas en una etapa posterior.

2. Utilizar una Licencia Abierta

- a) Determine qué derechos de propiedad intelectual existen en la información.
- b) Aplique una “Licencia Abierta” adecuada que licencie todos esos derechos y soporte la definición de apertura discutida en la sección anterior, en “Qué son los Datos Abiertos”. Puede consultar acerca de las licencias en los sitios de *Creative Commons* y *Open Data Commons*.
- c) *Nota:* Si tu no puede hacer esto, vuelva al paso 1 y pruebe con un *dataset* diferente.

3. **Poner a disposición la información-** en conjunto y en un formato útil. También puedes considerar otras alternativas de hacerla disponible, así como una API.

4. Hacerlos Visibles

Publíquelos en la web y si puede organice un catálogo central que enumere sus conjuntos de abiertos

Escoger las Bases de Datos

Escoger los conjuntos de datos que planea abrir es el primer paso, aunque recuerde que el proceso completo de abrir los datos es iterativo y puede regresar a este paso si encuentra problemas en un futuro.

Si usted ya conoce exactamente que conjunto(s) de dato(s) planea abrir, puede dirigirse a la siguiente sección. Sin embargo, en muchos casos, especialmente para grandes instituciones, escoger un conjunto de datos al cuál enfocarse es todo un reto. ¿Cómo se debería proceder en estos casos?

Crear esta lista puede ser un proceso rápido para identificar qué conjuntos de datos pueden ser abiertos para empezar. Ya habrá tiempo en etapas posteriores de revisar en detalle si cada conjunto de datos es adecuado para abrirse.

No hay requisitos para crear una lista completa de conjuntos de datos. El principal punto a tener en cuenta es pensar si es factible publicar todos sus datos (ya sea abiertamente o no) - vea la sección *¿De qué datos estamos hablando?*.

Pregunte a la comunidad

Recomendamos preguntar a la comunidad en primera instancia. Estas son las personas que tendrán acceso y utilizarán los datos, ya que es probable que tengan un mejor entendimiento de cuáles datos podrían ser valiosos.

1. Prepare una pequeña lista de los potenciales conjuntos de datos de los cuales desea tener comentarios. No es esencial que esta lista coincida con sus expectativas, la principal intención es obtener una idea de la demanda. Esto podría estar basado en otros catálogos de Datos Abiertos de otros países.
2. Solicite comentarios
3. Publique su petición en una página web. Asegúrese de que es posible acceder a dicha petición a través de su propia URL. De esta manera, cuando sea compartida a través de las redes sociales, la petición sea fácilmente encontrada.
4. Provea formas sencillas de enviar respuestas. Evite el requisito de registrarse previamente, puesto que reduce el número de respuestas.
5. Circule la petición a listas de correos relevantes, foros y participantes, señalando de nuevo la página web principal.
6. Realice un evento de consulta. Asegúrese de que se realice a una hora conveniente, donde el hombre de negocios promedio, los interesados en los datos y el representante del gobierno pueden asistir.
7. Pídale a un político que hable en nombre de su agencia. Es muy posible que Datos Abiertos sea parte de una política mayor de aumento del acceso a la información gubernamental.

Base del Costo

¿Cuánto dinero gastan las agencias en la recolección y mantenimiento de datos que poseen? Si gastan demasiado en un conjunto

de datos en particular, entonces es muy probable que otros deseen acceder a él.

Este argumento puede ser altamente susceptible a las preocupaciones del freeriding. La pregunta que necesita responder es: “¿Por qué otras personas deberían obtener información gratis si es tan cara?”. La respuesta es que los gastos son absorbidos por el sector público para realizar una función en particular. El costo de enviar los datos a un tercero, una vez que han sido recolectados, es aproximadamente nada. Por lo tanto, deberían cobrar nada.

Alivio de la Liberación

Algunas veces, en vez de decidir cuáles datos serían más valiosos, serían más útil el mirar cuáles datos son más fáciles de dar a manos públicas. Versiones pequeñas y fáciles pueden actuar como el catalizador de grandes cambios de comportamiento en las organizaciones.

Sin embargo, sea cuidadoso con esta aproximación. Quizás sea el caso que estas publicaciones sean de tan poco valor que nada puede construirse a partir de ellas. Si ocurre, la fé en todo el proyecto puede ser socavada.

Observe a sus Pares

Los Datos Abiertos es un movimiento en crecimiento. Es probable que haya mucha gente en su área que entiende lo que hacen otras áreas. Formule una lista en base a lo que esas agencias están haciendo.

Solicite una Licencia Abierta (Apertura Legal)

En la mayoría de las jurisdicciones existen derechos de propiedad intelectual sobre los datos, los cuales evita que terceros los usen, reutilicen y redistribuyan datos sin el permiso explícito. Aún en lugares donde la existencia de derechos es incierta, es importante aplicar una licencia simplemente en aras de la claridad. Por lo tanto, **si está planeando hacer disponible sus datos, debería ponerles una licencia** - y si desea que sus datos sean “libres”, esto es aún más importante.

¿Qué licencias puede usar? Recomendamos que para “abrir” datos use una de las licencias alineadas con la “Definición de Abierto” y marcarlas como adecuada para datos. Esta lista (además de las instrucciones para su uso) puede ser localizada en:

<http://opendefinition.org/licenses/>

Una breve guía (de una página) para utilizar una licencia para datos abiertos puede ser encontrada en el sitio Open Data Commons:

<http://opendatacommons.org/guide/>

Pon los datos a disposición (Apertura Técnica)

Los Datos Abiertos tienen que estar técnicamente como así también y legalmente abiertos. Específicamente, los datos tienen que estar disponibles *en crudo* en un formato legible por máquinas.

■ **Disponible**

La información no debe tener un precio mayor que un costo de reproducción razonable, preferentemente debe ser de descarga gratuita en Internet. El modelo de precios se consigue porque su agencia no debe entrar en costos al proveer la información para su uso.

■ **A Granel**

Los datos deberían estar disponibles como un conjunto completo. Si usted tiene un registro que se recoge en el estatuto, todo el registro debe estar disponible para su descarga. Una API web o servicio similar, también puede ser muy útil, pero no son sustitutos de acceso masivo.

■ **En un formato abierto, legible por máquinas**

La reutilización de los datos en poder del sector público no debe estar sujeto a restricciones de patentes. Más importante aún, es asegurarse de que usted está proporcionando los formatos legibles que permitan una mayor reutilización. Para ilustrar esto, considere las estadísticas publicadas como documentos PDF (*Portable Document Format*), a menudo usados para la impresión de alta calidad. Aunque estas estadísticas pueden ser leídas por los seres humanos, son muy difíciles de usarlas para un orde-

nador. Esto limita enormemente la capacidad para que otros puedan volver a utilizar esos datos.

Aquí hay algunas políticas que serán de gran beneficio:

- Manténgalo simple.
- Muévase rápido.
- Sea pragmático.

En particular es mejor brindar datos crudos ahora, que información perfecta en seis meses.

Hay muchas maneras de hacer que la información este disponible para terceros. La más natural en la era de Internet es la publicación en línea. Hay muchas variaciones para este modelo. En el más básico, las agencias disponibilizan la información a través de sus sitios web y un catálogo central direcciona a los visitantes a la fuente apropiada. Sin embargo, hay alternativas.

Cuando: término: la “conectividad” es limitada o el tamaño de los datos es demasiado grande, la distribución en otros formatos puede estar justificada. Esta sección también discutirá alternativas, que podrán accionar para mantener los precios muy bajos.

Métodos en línea

A través de tu sitio web existente

El sistema que será más familiar al equipo de contenidos de tu sitio web es el de proveer archivos para ser descargados desde páginas web. Así como provees acceso a documentos de discusión, los

archivos de datos serán muy felices de ser disponibilizados de esta manera.

Una dificultad propia de esta aproximación es que es muy difícil para terceros descubrir dónde encontrar información actualizada. Esta opción crea cierta carga para las personas que crean herramientas con tus datos.

A través de sitios de terceras partes.

Muchos repositorios se han convertido en centros de datos de campos específicos. Por ejemplo, *pachube.com* está diseñada para conectar personas con sensores, con aquellos que desean acceder a la información que producen. Sitios como *Infochimps.com* y *Talis.com* permiten a las agencias del sector público alojar grandes cantidades de datos gratuitamente.

Los sitios de terceros pueden ser muy útiles. La razón principal de esto es que ya se han agrupado juntos una comunidad de personas interesadas y otros conjuntos de datos. Cuando sus datos son parte de estas plataformas, se crea un tipo de interés compuesto positivo.

Las plataformas de datos ya proveen la infraestructura para soportar la demanda. Usualmente proveen análisis y usos de la información. Para las agencias del sector público suelen ser gratuitos.

Estas plataformas pueden tener dos tipos de costos. El primero es la independencia. Su agencia necesita ser capaz de ceder el control a otros. Esto es a menudo política y legalmente difícil. El segundo costo puede ser la apertura, es decir, que su plataforma de datos es

independiente de quién puede acceder a él. Los desarrolladores de software y los científicos usan muchos sistemas operativos, desde teléfonos inteligentes hasta superordenadores. Todos ellos deben ser capaces de acceder a los datos.

A través de servidores FTP

Un método menos de moda es el de proveer el acceso a los archivos a través del servicio FTP. Este puede ser un sistema adecuado si su audiencia es técnica, tanto como desarrolladores de software como científicos. El sistema FTP trabaja en lugar del HTTP, pero específicamente se utiliza para la transferencia de archivos.

El FTP cayó en desgracia. En lugar de proveer un sitio web, mirar un FTP es como mirar entre carpetas de una computadora. Entonces, aunque está en las condiciones para su fin, las compañías de desarrollo web tienen mucha menos capacidad para cobrar por la personalización.

Como archivo Torrent

“*BitTorrent*” es un sistema que se ha vuelto familiar para los encargados de crear políticas, y esto se debe a su asociación con las infracciones de copyright. BitTorrent utiliza archivos llamados Torrents, que funcionan separando el costo de distribución de los archivos entre toda la gente que accede a esos archivos. En lugar de servidores sobrecargados, el suministro se incrementa con un aumento de demandas. Esta es la razón por la que este sistema es tan exitoso para compartir películas. Se trata de una manera eficiente de distribuir grandes volúmenes de datos.

Como un archivo API

Los datos pueden ser publicados a través de un *Programa de Interfaz de Aplicaciones* (API). Estas interfaces se volvieron muy populares. Permiten a los programadores seleccionar porciones específicas de información, en lugar de proveerla toda en grandes cantidades en un archivo de gran tamaño. Las APIs usualmente están conectadas a una base de datos que es actualizada en tiempo real. Esto significa que disponibilizar la información a través de una API garantiza que esté actualizada.

Publicar datos crudos en grandes cantidades debería ser la primera preocupación de todas las iniciativas de datos abiertos. Proveer de una API requiere cierto costo.

1. El precio. Ellos requieren mucho más desarrollo y mantenimiento que ofrecer solo los archivos.
2. Las expectativas. Con el fin de fomentar una comunidad de usuarios detrás del sistema, es importante proporcionar seguridad. Cuando las cosas van mal, se espera que absorbas los costos para solucionarlo.

El acceso a datos en grandes cantidades asegura que:

1. No hay dependencia del proveedor original de la información, es decir, si la situación cambia por una reestructuración o una modificación presupuestaria, la información todavía estará disponible.

2. Cualquiera puede obtener una copia y redistribuirla. Esto reduce los costos de distribución desde la agencia y significa que no hay ningún punto librado al azar.
3. Otros pueden desarrollar sus servicios usando los datos, porque tiene seguridad que los datos no van a ser robados.

Proveer datos en masa permite a otros usar esa información para propósitos distintos de sus originales. Por ejemplo permite que los datos sean convertidos a nuevos formatos, enlazados con otros recursos o versionados y archivados en múltiples lugares. Mientras que la última versión de la información puede ser disponibilizada a través de una API, datos crudos pueden ser disponibilizados en masa durante intervalos de tiempo regulares.

Por ejemplo, el servicio estadístico *Eurostat* <http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home/> tiene un sitio de descargas por volumen que ofrece más de 4000 archivos. Lo actualizan dos veces al día, ofrece datos en formato (TSV) Valores separados por Tabulador *Tab-Separated Values*, e incluye documentación sobre el sitio de descarga, así como los archivos de datos.

Otro ejemplo es el Catálogo de Datos del Distrito de Columbia <http://octo.dc.gov/DC/OCTO/>, que permite que los datos sean descargados en formato CSV y XLS, además de las transmisiones en vivo de los datos.

Haga Datos Detectables

Los Datos Abiertos no son nada sin los usuarios. Tiene que ser capaz de asegurarse de que las personas pueden encontrar el material de origen. Esta sección cubrirá diferentes enfoques.

Lo más importante es proporcionar un espacio neutral que puede superar tanto la política y los ciclos presupuestarios futuros. Fronteras jurisdiccionales, ya sea sectorial o geográfica, pueden hacer difícil la cooperación. Sin embargo, hay beneficios significativos cuando se unen fuerzas. Cuanto más fácil sea para los de afuera descubrir los datos, más rápido se construirán herramientas nuevas y útiles.

Las Herramientas Existentes

Hay una serie de herramientas que están en la web diseñadas específicamente para hacer a los datos más detectables.

Uno de los más prominentes es el *DataHub*, un catálogo y almacén de bases de datos de todo el mundo. El sitio facilita a los individuos y a las organizaciones publicar materiales y a los usuarios de los datos encontrar el material que necesitan.

Además, hay docenas de catálogos específicos para diferentes sectores y lugares. Comunidades específicas han creado sistemas de catálogos para sus archivos, ya que a menudo los datos requieren de publicación

Para el Gobierno

Como se ha comprobado, la práctica ortodoxa es que una agencia líder para cree un catálogo de datos del gobierno. Cuando se establece un catálogo, se trata de crear una estructura que permite a muchos departamentos seguir fácilmente su propia información actualizada.

Resista la tentación de desarrollar un software para generar el catálogo. Existen soluciones de software libre y de código abierto (como *CKAN*) que han sido adoptadas por muchos gobiernos. Invertir en otra plataforma puede que no sea necesario.

Hay algunas cosas que la mayoría de los catálogos de apertura de datos no tienen. Tu programa debe considerar lo siguiente:

- Proporcionar un canal para que los privados y sectores de la comunidad añadan sus datos. Puede valer la pena pensar en el catálogo como el catálogo de la región, en lugar de el catálogo del gobierno regional.
- Facilitar la mejora de la información permitiendo la catalogación de derivados de los conjuntos de datos. Por ejemplo, alguien podría geocodificar direcciones y compartir los resultados con todos. Si sólo permites versiones individuales de conjuntos de datos, estas mejoras podrían permanecer escondidas.
- Tolera que tu información aparezca en otros lugares. Es decir, el contenido probablemente será duplicado por las comunidades de interés. Si tienes datos de monitoreo del nivel de agua en un río, entonces tus datos podrán aparecer en catálogos para hidrólogos.

- Asegure que el acceso sea equitativo. Evite crear niveles privilegiados de acceso para funcionarios o investigadores, puesto que afectará la participación y el involucramiento de la comunidad.

Para la Sociedad Civil

Estar dispuesto a crear un catálogo complementario para los datos no oficiales.

Es muy raro que los gobiernos se asocien con fuentes no oficiales o sin autoridad. Los funcionarios han invertido mucho para asegurarse de que no existan implicaciones políticas u otro tipo de daños asociados con el uso inadecuado de los datos.

Más aun, los gobiernos no suelen estar dispuestos a soportar actividades que mezclen su información con empresas. Los gobiernos son escépticos, y con razón, a los motivos con fines de lucro. Por lo tanto, un catálogo independiente para comunidades, empresas y otros puede estar garantizado.

Capítulo 5

Ya he Publicado Cierta Información, ¿y ahora qué?

Ya hemos visto cómo hacer que la información del gobierno sea legal y técnicamente reutilizable. El siguiente paso es motivar a que otros hagan uso de esa información.

Esta sección sería como algo adicional que puede promover la reutilización de datos.

¡Avísenle al Mundo!

En primer lugar, asegúrese de promover el hecho de que usted ha puesto en marcha una campaña de promoción de Datos Abiertos en su área de responsabilidad.

Si publica una serie de datasets, definitivamente vale la pena dedicarle tiempo a asegurarse que la gente sepa (o que al menos pueda averiguar) que lo ha hecho

Además de comunicados de prensa, anuncios en tu sitio web, etc., podrías considerar lo siguiente:

- Contactando a organizaciones importantes o individuos que trabajan o están interesados en el tema.
- Contactando a bases de datos o grupos de redes sociales que sean relevantes.
- Contactar directamente con usuarios potenciales, de los cuales sepamos su interés en los datos.

Comprendiendo tu Audiencia

Como toda la comunicación pública, todo lo relacionado a la comunidad de datos, debe estar dirigida al público correcto. Ya que el mensaje puede ser desperdiciado si va dirigido a una área equivocada.

Las comunidades digitales están siempre bien dispuestas a compartir nueva información, a pesar de que ésta se consume rápidamente. Escribe mensajes que puedan ser “escaneados”, en vez de aquellos que requieran ser examinados en profundidad.

Los miembros de la comunidad tecnológica, son más renuentes que el público en general en usar MS-Windows. Esto significa que usted no debe guardar documentos en formatos MS Office que puedan ser leídos en línea. Hay dos razones para esto:

- La primera es que esos documentos serán menos accesibles. Además del documento que usted ve en su pantalla, los lectores pueden ver una copia imperfecta como alternativa.
- Segundo, tu agencia envía un mensaje implícito diciendo que no estás dispuesto a avanzar hacia los desarrolladores. En su lugar, deberías demostrar que tú estás esperando que la comunidad tecnológica se acerque a tí.

Publica tu Material en los Sitios Web de Terceros

Muchos blogs le han dedicado un gran espacio a temas especializados. Quizás valdría la pena incluir un artículo sobre tu iniciativa en sus sitios. Esto podría ser un beneficio mutuo. Tú recibes mayor interés y ellos una publicación gratis en esa área específica.

Haciendo que tus Comunicaciones Sean más Amigables con las Redes Sociales

No es realista esperar que agentes del gobierno esperen largos períodos de tiempo relacionandose con los medios. Sin embargo, hay muchas cosas que puedes hacer para asegurarte de que tu contenido sea fácil de compartir entre usuarios técnicos. Algunos consejos:

- Proporcione páginas únicas para cada pieza de contenido:
Cuando un mensaje es compartido con otros, el destinatario buscará rápidamente el contenido relevante.
- Evite hacer que la gente deba descargar sus notas de prensa:
Las publicaciones de prensa están bien. Son mensajes concisos sobre un punto en particular. Sin embargo, si requieres que las personas se descarguen el contenido y éste se abra fuera del sitio, menos personas lo leerán. Las personas son poco propensas a hacer un clic para descargar la información.
- Considere el uso de una Licencia Abierta para su contenido:
Aparte de brindar asegurar a las personas que lo deseen que está permitido compartir tu contenido, mandas un mensaje de que tu departamento entiende lo que es ser “abierto”. Esto deja una impresión mucho más significativa a los defensores de los datos abiertos que cualquier frase de tu gacetilla de prensa.

Redes Sociales

No es eficiente para agencias con poco presupuesto dedicar horas en sitios de redes sociales. La manera más significativa para que tu voz sea escuchada en estos foros es asegurándote que las publicaciones de tu blog se puedan compartir fácilmente. Eso significa, antes de leer la próxima sección asegúrate de haber leído la última. Con eso en mente, algunas sugerencias:

- Foros de Discusión:
Twitter ha surgido como la plataforma predilecta para diseminar información de forma muy rápida. Cualquier cosa etique-

tada con *#OpenData* o *#DatosAbiertos* podrá ser vista inmediatamente por miles de personas.

Linkedin tiene una gran selección de grupos que están dirigidos a los datos abiertos.

Si bien *Facebook* es excelente para una audiencia general, no ha recibido tanta atención dentro de la comunidad de los Datos Abiertos.

- Los agregadores de enlace:

Envía tu contenido al equivalente de un cable de noticias para geeks. *Reddit*¹ y *Hacker News*² son actualmente los dos más grandes. Le siguen *Slashdot*³ y *Digg*⁴, que también son dos herramientas útiles en esta área.

Estos sitios tienen una tendencia a manejar un tráfico de material interesante. También están muy centradas en las áreas temáticas.

¹Reddit: <http://www.reddit.com/>

²Hacker News: <http://news.ycombinator.com/>

³Slashdot: <http://slashdot.org/>

⁴Digg: <http://www.digg.com/>

Reunir personas en una habitación: Desconferencias, Encuentros y Barcamps

Los eventos cara a cara pueden ser una vía muy afectiva para fomentar el uso de los datos. Razones para considerar en un evento:

- Conociendo más acerca de reutilizadores potenciales.
- Conociendo más acerca de la demanda de diferentes datasets
- Descubriendo mas acerca de como las personas quieren rehusar sus datos
- Habilitando a reutilizadores potenciales a conocer más acerca de los datos que tu tienes
- Posibilitando que reutilizadores potenciales se conozcan entre sí (para que puedan colaborar)
- Exponiendo tu información a una audiencia más amplia (por ejemplo, mediante posts en blogs o cobertura de prensa que el evento puede ayudar a generar)

Hay también un montón de diferentes maneras de ejecutar eventos, y diferentes tipos de eventos, depende del objetivo que se pretenda alcanzar. Así como los modelos más tradicionales de la conferencia, que incluirá cosas como conversaciones formales preparadas, presentaciones y demostraciones, también hay varios tipos de eventos impulsados por los participantes, donde aquellos que podrían presentarse:

- Guía para definir la agenda del evento

- Presentándose a si mismos, hablando acerca de lo que les interesa, y en lo que están trabajando, sobre una base ad hoc.
- Brinda micro-presentaciones improvisadas sobre algo que estás trabajando
- Liderar sesiones sobre algún tema en el que estén interesados.

Hay un montón de documentación en línea sobre cómo ejecutar este tipo de eventos, que se pueden encontrar mediante la búsqueda de cosas como: *“desconferencia”*, *“BarCamp”*, *“quedada”*, *“speed-geek”*, *“charla relámpago”*, y así sucesivamente. Usted también puede encontrar que vale la pena ponerse en contacto con personas que ya han ejecutado este tipo de eventos en otros países, y lo más probable es que se encuentren dispuestos a ayudarlo y asesorarlo en su evento. Puede ser útil asociarse con otra organización (por ejemplo, una organización de la sociedad civil, una organización de noticias o de una institución educativa) para ampliar su base de participantes y aumentar su exposición.

¡Haciendo cosas! Hackatons, premios, prototipos

La estructura de estos concursos es que una serie de conjuntos de datos serán liberados y los programadores tendrán un breve espacio de tiempo - que va desde tan sólo 48 horas hasta un par de semanas- para desarrollar aplicaciones utilizando esos datos. Entonces se le otorgará un premio a la mejor aplicación. Este tipo de competencias se han celebrado en varios países incluyendo el Reino Unido, EE.UU., Noruega, Australia, España, Dinamarca y Finlandia.

Ejemplos de competencias

Show us a better way fue la primera competición del estilo en el mundo. Bajo iniciativa del grupo de trabajo Gobierno del Reino Unido de nombre “El poder de la información” coordinado por el Jefe de Gabinete de Ministro Tom Watson en marzo de 2008. Esta competición se preguntaba “¿qué crearías con información pública?” y se abrió a programas de todo el mundo con un tentador premio de 80 mil libras para las mejores cinco aplicaciones.

Apps for democracy una de los primeros concursos de los Estados Unidos, fue lanzado en octubre de 2008 por Vivek Kundra, en ese momento Jefe de la Oficina de Tecnología del Gobierno del Distrito de Columbia. Kundra había desarrollado el catálogo de Datos del DC, <http://data.octo.dc.gov/>, que incluye datos de crímenes en tiempo real, resultados de exámenes en las escuelas e indicadores de pobreza. Fue al momento el catálogo de datos locales más exhaustivo del mundo. El desafío era hacerlo útil para los ciudadanos,

visitantes, empresarios y miembros de del Gobierno de Washington DC.

La solución creativa fue llevar a cabo el concurso Apps For Democracy. La estrategia consistió en pedir a la gente que utilice los datos para construir aplicaciones, datos pertenecientes al catálogo recién lanzado. Se incluyó una presentación en línea para aplicaciones, muchos premios pequeños en lugar de unos pocos grandes, con distintas categorías diferentes, así como un premio por “Elección del Público”. El concurso estuvo abierto durante 30 días y le costó al gobierno de DC la suma de 50.000 dólares. Por su parte, fueron desarrollados un total de 47 aplicaciones para iPhone, Facebook y web, con un valor estimado para la economía local de más de 2.600.000 USD .

El desafío **Abre Datos 2010** Llevado a cabo en España en abril de 2010, este concurso invita a desarrolladores a crear aplicaciones abiertas haciendo uso de datos públicos en sólo 48 horas. La competición tuvo 29 equipos de participantes que desarrollaron aplicaciones que incluyeron: un programa para acceder a información sobre el tráfico desde teléfonos celulares en el País Vasco, y otro para acceder a datos sobre colectivos y paradas en Madrid. Ambos ganaron el primer y el segundo premio respectivamente.

Netskap 2.0 En abril de 2010 el Ministerio para la Administración de Gobierno llevó a cabo Nettskap 2.0. Desarrolladores noruegos -compañías, agencias públicas e individuos- fueron desafiados a crear proyectos con base en internet en las áreas de desarrollo de servicios, procesos de trabajo eficientes e incremento de la participación democrática. El uso de datos de Gobierno fue promovido explícitamente. Aunque la fecha límite para presentar los proyectos

fue solamente un mes después, el 9 de mayo, el Ministro Rigmor Aasrud dijo que la respuesta fue “abrumadora”. Se recibieron el total de 137 aplicaciones, no menos del 90 % de ellas fueron creadas reutilizando datos públicos. Un total de 2,5 millones de dólares noruegos fueron distribuidos entre los 17 ganadores; mientras que el valor acumulado por las 137 aplicaciones fue de 28,4 millones.

Australia Mashup. El grupo de trabajo que coordina la iniciativa de Gobierno 2.0 de Australia invitó a los ciudadanos a demostrar por qué el acceso abierto a la información del Gobierno es positiva para la economía y el desarrollo social del país. El concurso duró del 7 de octubre al 13 de noviembre de 2009. El equipo de trabajo publicó algunos datos bajo una licencia abierta y en formatos reutilizables. Las 82 aplicaciones que entraron en el concurso evidencian la gran cantidad de nuevas e innovadoras aplicaciones que pueden crearse a partir de la publicación de datos de gobierno en formatos abiertos.

Desarrollando América Latina es la hackathon colaborativa más grande de la región latinoamericana donde se crean aplicaciones con datos abiertos, y cuyos objetivos son generar soluciones para problemas sociales como salud, educación, seguridad ciudadana entre otros. Desarrollando América Latina es una iniciativa regional que reúne ciudadanos, instituciones y gobiernos para la creación de soluciones innovadoras y escalables para problemas sociales comunes de la región. La iniciativa impulsa junto a cada país, la apertura de los datos y el uso y reutilización de éstos, durante un maratón de desarrollo de 36 horas, con cientos de desarrolladores, diseñadores gráficos, emprendedores, periodistas y representantes del sector privado y público. El evento ya se encuentra presente en 8 países,

y en su segunda versión, apunta a entregar soluciones de alto impacto a problemas en salud, educación, seguridad ciudadana, gasto estatal, pobreza, medioambiente entre otros. La hackathon del año 2012 se realizó el 1 y 2 de Diciembre en Argentina, Bolivia, Brasil, Costa Rica, Chile, México, Perú, y Uruguay. Las aplicaciones que se crearon en la hackathon apuntan a mejorar justamente estos problemas sociales. Para crear estas aplicaciones, los participantes tuvieron a su disposición datos abiertos de los gobiernos, instituciones nacionales, organizaciones de la sociedad civil, y más.

Conferencias, Barcamps, Hackatons

Una de las maneras más eficaces para las Organizaciones de la Sociedad Civil (OSC / ONG) para demostrar a los gobiernos el valor de la apertura de sus bases de datos es mostrar las múltiples formas en que la información puede ser administrada para lograr un beneficio social y económico. Las OSC que promueven la reutilización han sido fundamentales en los países donde se han producido avances en la política y la ley para asegurar que los conjuntos de datos sean a la vez técnica y jurídicamente abiertos.

Las típicas actividades que se desarrollan como parte de estas iniciativas normalmente incluyen competencias, “conferencias de Datos Abiertos de Gobierno”, “anticonferencias”, workshops y hackatons. Estas actividades suelen estar organizadas por la comunidad de usuarios con datos que ya han sido publicados proactivamente u obtenidos haciendo peticiones de acceso a la información. En otros casos, los defensores de la sociedad civil han trabajado con miembros de gobierno progresistas para asegurar nuevas publica-

ciones de datos que puedan ser usados por programadores para crear aplicaciones innovadoras.

Glosario

Anonimización: El proceso de adaptación de datos de manera que los individuos no pueden ser identificados a partir de ella.

API: Application Programming Interface

Una manera en que los programas informáticos se hablan entre ellos. Puede ser entendido en términos de cómo un programador envía instrucciones entre los programas.

AR: Ver “Registro de información de recurso”

Attribution License: Una licencia que requiere que la fuente original del material autorizado sea citada (atribuida).

BitTorrent: BitTorrent es un protocolo para distribuir la banda ancha que permite transferir archivos muy grandes entre computadoras. En lugar de descargar un archivo desde una fuente específica, BitTorrent permite la descarga entre pares.

Connectivity: Conectividad se relaciona con la habilidad de las comunidades para comunicarse a Internet, especialmente a la World Wide Web.

Copyright: El derecho de los creadores de las obras creativas para restringir el uso de los demás de esas obras. Un dueño de los derechos de autor tiene derecho a determinar cómo otros pueden usar ese trabajo.

DAP: Ver: término: “Protocolo de Acceso a Datos”

Data Access Protocol: Un sistema que le permite a terceros obtener acceso a bases de datos sin sobrecargar ninguno de los sistemas.

Data protection legislation: La legislación sobre protección de datos, no se basa en proteger data, pero sí en proteger el derecho de los ciudadanos de poder vivir sin miedo en cuanto a que su información privada no se hará pública. La ley protege la privacidad (ya sea información sobre el estado económico de una persona, su salud o posición política), y otros derechos como el derecho a una libertad de circulación. Por ejemplo, en Finlandia un sistema de tarjeta de transporte se utiliza para registrar todas las veces que la tarjeta se muestra en la máquina lectora en diferentes líneas de transporte público. Esto generó un debate desde la perspectiva de la libertad de circulación y la tarjeta de viaje de recolección de datos fue abandonada sobre la base de la legislación sobre protección de datos.

Database rights: Un derecho que prohíbe a otros extraer y utilizar contenido de una base de datos. Existe principalmente en jurisdicciones europeas.

EU: Unión Europea.

EU PSI Directive:

IAR: Ver “Registro de información de recurso”

Information Asset Register:

IARs son registrados específicamente para capturar y organizar los meta-datos acerca de las enormes cantidades de información en poder de los departamentos y organismos gubernamentales. Un IAR integral incluye bases de datos, antiguos conjunto de archivos, últimos archivos electrónicos, recopilaciones de datos, la investigación y así sucesivamente.

El término: Directiva de la Unión Europea sobre Información del Sector Público reconoce la importancia de los registros más atractivos para posibles reutilizadores de información del sector público. Requiere que los estados miembro provean listas, portales o algo similar. Dice:

Tools that help potential re-users to find documents available for re-use and the conditions for re-use can facilitate considerably the cross-border use of public sector documents. Member States should therefore ensure that practical arrangements are in place that help re-users in their search for documents available for reuse. Assets lists, accessible preferably online, of main documents (documents that are extensively re-used or that have the potential to be extensively re-used), and portal sites that are linked to decentralised assets lists are examples of such practical arrangements.

Las IARs pueden ser desarrollados de diferentes maneras. Los departamentos de Gobierno pueden desarrollar sus propias IARs y éstas pueden asociarse a IARs nacionales. Las IARs pueden

incluir información del sector público que no ha sido -y quizás no sea- publicada proactivamente. Por lo tanto permite a miembros del público identificar información que existe y puede ser requerida.

Para que el público use estos IARs, es importante que cada registro de información sea lo más completo posible para tener seguridad de que los documentos puedan encontrarse. Que algunos registros estén incompletos es un problema porque crea cierto grado de incertidumbre que puede desincentivar a algunas personas de usarlos para buscar información.

Es muy importante que los metadatos en el IAR sean exhaustivos de manera que los motores de búsqueda funcionen de manera efectiva. En el espíritu de las iniciativas de Datos Abiertos de Gobierno, éste debería disponibilizar sus IARs al público en general como datos crudos bajo licencia abierta, así hackers cívicos pueden usarlos, por ejemplo para construir motores de búsqueda o interfaces.

Intellectual property rights: Monopolios otorgados a personas para las creaciones intelectuales

IP rights: Ver: término: “Derechos de Propiedad Intelectual”

Machine-readable: Los formatos leíbles por máquina son aquellos que sus datos pueden ser extraídos por programas de computadoras fácilmente. Los documentos PDF no son legibles a las máquinas. Las computadoras pueden el texto muy bien, pero tienen grandes dificultades entendiendo el contexto que los rodea.

Open Data: Los datos abiertos pueden ser usados para cualquier propósito. Para más detalles: <http://www.opendefinition.org/>.

Open Government Data: Del término: “Open Data” “Datos Abiertos” producidos por el gobierno. Esto es generalmente aceptado por ser datos recopilados durante el curso de los negocios, como aquellas actividades habituales que no identifican a los individuos o sensibilidad comercial. Abrir datos del gobierno es un subconjunto del término: “Información del Sector público”, el cual es más amplio en su alcance. Ver: <http://opengovernmentdata.org> para más detalles.

Open standards: En general se entiende estándares técnicos libres de restricciones de licencias. También puede ser interpretado en el sentido los estándares que se desarrollan de manera independiente del proveedor.

PSI: Ver: término: “Información del Sector Público”.

Public domain: No existen derechos de autor sobre el trabajo.No existen en todas las jurisdicciones.

Public Sector Information: Información recolectada o controlada por el sector público.

Re-use: Uso de contenido por fuera de su intención original.

Share-alike License: Una licencia que requiere que los usuarios de un trabajo provean el contenido bajo condiciones similares o iguales a las del trabajo original.

Tab-separated values: Los valores-separados por tabulador (TSV) son una muy común formato de archivo de texto para compartir datos tabulares. Este formato es extremadamente simple y altamente leible por la máquina-machine-readable

Web API: Un “API” que está diseñado para funcionar en Internet.

Formatos de Archivos

Un Repaso de los Formatos de Archivos

- **JSON**

JSON es un formato de archivo que es muy fácil de leer por cualquier lenguaje de programación. Su simplicidad significa que es generalmente fácil de procesar para las computadoras a comparación de otros, como lo es XML.

- **XML**

XML es un formato ampliamente utilizado para el intercambio de datos debido a que ofrece buenas oportunidades de mantener la estructura de los datos y la forma en que los archivos son construidos, y permite a los desarrolladores escribir partes de la documentación con los datos sin interferir con la lectura de ellos.

- **RDF**

Un formato recomendado por W3C llamado RDF permite representar datos en una forma que facilita la combinación de información de diferentes fuentes. Los datos RDF pueden ser al-

macenados en XML y JSON, entre otras serializaciones. RDF promueve el uso de URLs como identificadores, lo que brinda una manera conveniente de interconectar iniciativas de “Datos Abiertos” existentes. RDF todavía no se ha generalizado, pero es una tendencia entre iniciativas de Gobierno Abierto, incluso las iniciativas de Linked Data de los gobiernos británico y español. El inventor de la Web, *Tim Berners-Lee*, propuso recientemente un esquema de puntuación de cinco estrellas que incluye *Linked Data* en formato RDF como una meta a alcanzar en las iniciativas de Datos Abiertos.

■ **Hojas de Cálculo**

Muchas autoridades tienen información almacenada en las hojas de cálculo, por ejemplo Microsoft Excel. Estos datos a menudo pueden ser utilizados inmediatamente con las descripciones correctas, de lo que significan las distintas columnas.

Sin embargo, en algunos casos pueden haber macros y fórmulas en las hojas de cálculo, las cuales pueden ser más incómodas para manipular. Por tanto, es recomendable documentar este tipo de cálculos junto a la hoja de cálculo, ya que generalmente es más accesible al leer por los usuarios.

■ **Archivos Separado por Comas (CSV)**

Los archivos CSV pueden ser un formato útil, debido a que son compactos y por lo tanto adecuados para transferir grandes conjuntos de datos con la misma estructura. Sin embargo, el formato es tan espartano que los datos son frecuentemente inservibles sin una documentación que los explique, ya que puede ser casi imposible adivinar el significado de las diferentes columnas. Por tanto, es particularmente importante para los formatos se-

parados por comas que la documentación de los campos individuales sea precisa.

Además, es esencial que la estructura del archivo sea respetada. La omisión de un solo campo puede alterar la lectura de todos los datos restantes en el archivo sin ninguna posibilidad real de rectificarla, porque no se puede determinar cómo deben ser interpretados los datos restantes.

■ **Documento de Texto**

Documentos clásicos en formatos como Word, ODF, OOXML, o PDF, pueden ser suficientes para mostrar ciertos tipos de datos, por ejemplo, listas de correo o equivalentes. Puede que sea barato para exponerlos, siempre que ese sea el formato en que los datos se han creado. El formato no da soporte para mantener la estructura coherente, lo cual significa que a veces es difícil introducir datos por medios automatizados. Asegúrese de utilizar las plantillas como base de documentos en las que se mostrarán los datos para su reutilización, por lo que al menos sea posible extraer información de los documentos.

También se puede apoyar el uso posterior de los datos que utilice el marcador de la tipografía tanto como sea posible para que sea más fácil para una máquina distinguir las partidas (de cualquier tipo) a partir del contenido y así sucesivamente. En general se recomienda no exhibir en formato de documentos de texto, si hay datos en un formato diferente.

■ **Texto Plano**

Los documentos de texto plano (.txt) son muy fáciles de leer en los ordenadores. Por lo general, no incluyen metadatos estructurales, lo que significa que los desarrolladores necesitan crear

un programa de análisis que pueda interpretar cada documento tal y como aparece.

Algunos problemas pueden ser causados al cambiar archivos de texto plano entre distintos sistemas operativos. MS Windows, Mac OSX y otras variantes de Unix tienen su propia forma de decirle a la computadora que se ha llegado al final de la línea.

- **Imagen Escaneada**

Probablemente sea la forma menos adecuada para la mayoría de los datos, pero en ambos, el formato TIFF y JPEG 2000 (al menos puede describir con la documentación de lo que hay en la imagen) son adecuados para marcar la imagen de un documento con contenido de texto completo del documento. Puede ser relevante mostrar datos como imágenes de aquellos datos que no nacen por la vía electrónica -un ejemplo obvio son los registros antiguos de la iglesia y material de archivo- y una imagen es mejor que nada.

- **Formatos Proprietarios**

Algunos sistemas tienen sus propios formatos de datos que pueden guardar o exportar datos. A veces puede ser suficiente para exponer los datos en un formato - sobre todo si se espera que el uso de más estaría en un sistema similar del que vienen. Debe estar siempre indicado sobre donde obtener más información sobre estos formatos, por ejemplo, al proporcionar un enlace a la página web del proveedor. En general se recomienda mostrar los datos en formato libre, cuando sea factible.

- **HTML**

Hoy en día, muchos datos se encuentran disponibles en varios sitios en formato HTML. Esto sería suficiente, si se trata de da-

tos estables y de alcance limitado. En algunos casos, sería preferible tener los datos de una forma fácil de descargar y manipular, pero como es barato y fácil de hacer referencia a una página en un sitio web, podría ser un buen punto de partida para mostrar datos en una pantalla.

Generalmente, sería más apropiado el uso de tablas en documentos HTML para almacenar datos y, es importante que los distintos campos de datos se muestren y se provean identificaciones, que hacen más fácil encontrar y manipular los datos. *Yahoo* ha desarrollado una herramienta (<http://developer.yahoo.com/yql/>) que puede extraer información estructurada de un sitio web, y ciertas herramientas que pueden hacer mucho más con los datos siempre que se encuentren cuidadosamente etiquetados.

Archivos de Formatos Abiertos

Incluso si la información se proporciona en formato electrónico, formato legible por máquina, y en detalle, puede haber problemas relacionados con el formato del archivo en sí.

Los formatos en los cuales la información es publicada - en otras palabras, la base digital en la cual la información es almacenada - puede ser "*abierto*" o "*cerrado*". Un formato abierto es aquel donde las especificaciones del software están disponibles para cualquier persona, de forma gratuita, así cualquiera puede usar dichas especificaciones en su propio software sin ninguna limitación en su reutilización que fuere impuesta por derechos de propiedad intelectual.

Si el formato del archivo es “cerrado”, esto puede ser debido a que el formato es propietario y sus especificaciones no están disponibles públicamente, o porque el formato es propietario y aunque las especificaciones se han hecho públicas, su reutilización es limitada. Si la información es liberada en un formato de archivo cerrado, esto puede causar grandes obstáculos para reutilizar la información codificada en él, forzando a aquellos que deseen usar la información a comprar software innecesario.

La ventaja de los archivos de formatos abiertos, es que permiten a los desarrolladores producir varios paquetes de software y servicios utilizando esos formatos. Esto entonces reduce al mínimo los obstáculos para la reutilización de la información que contienen.

El uso de formatos de archivo con propiedad, para el que la especificación no está disponible públicamente, puede crear dependencia de software de terceros o de los titulares de licencias de los formatos de archivos. En el peor de los casos, esto puede significar que la información sólo se puede leer con ciertos softwares especiales, que pueden ser muy caros, o que pueden quedar obsoletos.

La preferencia del término *Gobierno de Datos Abiertos*, es que la información se publicará en formatos de archivo abiertos, los cuales son de lectura mecánica.

Ejemplo: datos del tráfico en Reino Unido

Andrew Nicholson es un desarrollador de software que estuvo involucrado en una campaña (la cual es un éxito) contra la construcción de una nueva carretera, la carretera de circunvalación al este de Westbury, en el Reino Unido. Andrew estaba interesado en acceder y utilizar los datos de tráfico que se estaban utilizando para

justificar las propuestas. Se las arregló para obtener algunos de los datos relevantes a través de un requerimiento de libertad de información, pero el gobierno local proporcionó los datos en un formato propietario que sólo se pueden leer con el software producido por una empresa llamada Saturno, que se especializa en el modelado de tráfico y las previsiones.

No existe ninguna versión disponible del software para usarlo en modo “solo lectura”, por lo que Andrew grupo no tuvo más remedio que adquirir una licencia de software, eventualmente pagando £500 al hacer uso de un descuento educativo. Los principales paquetes de software incluidos en la lista de precios de abril de 2010 de Saturno comienzan en £13.000, un precio que está fuera del alcance de la mayoría de los ciudadanos comunes.

Aunque las leyes de acceso a la información no da un derecho a acceder a información de formatos abiertos, las iniciativas de *Datos Abiertos del Gobierno* (OGD) están comenzando a ir acompañadas de una política que estipula que la información oficial debe estar disponible en formatos abiertos. Ha sido el Gobierno de Obama el responsable de establecer el estándar de oro, cuando lanzó la *Directiva de Gobierno Abierto*, emitida en Diciembre de 2009, la cual estableció:

En la medida de lo posible y con sujeción a las restricciones vigentes, las agencias deben publicar información en línea en un formato abierto que se pueda recuperar, descargar, indexar y buscar por las aplicaciones de uso común de búsqueda web. Un formato abierto es una plataforma que es independiente de la máquina, fácil de leer, y puesto a disposición del público sin restricciones que impidan la reutilización de esa información.

¿Cómo uso un Formato Determinado?

Cuando una autoridad debe mostrar datos nuevos - datos que no han sido expuestos antes - se debe elegir el formato que proporciona el mejor equilibrio entre el costo y la idoneidad para el propósito. Para cada formato, hay algunas cosas que usted debe tener en cuenta, y esta sección tiene como objetivo explicarlas.

Esta sección sólo se centra en cómo las superficies de corte están mejor organizadas de modo que las máquinas puedan acceder a ellas directamente.

En otros lugares puede encontrar asesoramiento y orientación acerca de cómo deben ser diseñadas las soluciones web y los sitios de internet.

Servicios Web

Cuando los datos cambian frecuentemente, y el tamaño de cada descarga es limitado, es muy importante exponer la información a través de *web services*. Hay muchas maneras de crear web servicios, pero algunos de más usados son SOAP y REST. Generalmente SOAP sobre REST. Los servicios REST son fáciles de desarrollar y entender, por eso es un standard muy usado.

Base de Datos

Como los servicios web, las bases de datos permiten el acceso directo a los datos dinámicamente. Las bases de datos tienen la ventaja de extraer aquella información en la cual estén interesados.

Hay algunas preocupaciones de seguridad en cuanto a permitir extracciones remotas de las bases de datos, y el acceso a las bases de datos es útil si su estructura y la importancia de las tablas y campos individuales están bien documentados. Generalmente es relativamente simple y no tiene costo crear web services que expongan datos de una base de datos, esta puede ser una manera fácil de lidiar con las problemáticas de seguridad.

¿Qué derechos legales (IP) hay en las Bases de datos?

Cuando se habla sobre bases de datos, en primer lugar hay que distinguir entre la estructura y el contenido de una base de datos (cuando utilizamos el término “datos”, nos referimos al contenido de la propia base de datos). Los elementos estructurales incluyen cosas como los nombres de campo y un modelo de organización de datos en esos campos y la relación entre ellos.

En muchas jurisdicciones, es probable que los elementos estructurales de una base de datos serán cubiertos por derechos de autor (esto dependerá del nivel de “creatividad” involucrados en la creación de esta estructura).

Sin embargo, aquí estamos particularmente interesados en los datos. Cuando hablamos de “datos” tenemos que ser un poco cuidadosos, porque la palabra no es muy precisa: “datos” puede significar unos cuantos artículos, o incluso un solo elemento (por ejemplo, un único registro bibliográfico, etc) o “datos” puede significar una gran colección (por ejemplo, todo el material en la base de datos).

Para evitar confusiones, nos reservamos el término “contenido” para referirnos a los elementos individuales, y “datos” para referirse a un conjunto.

A diferencia de los materiales, los textos, la música o el cine, la situación jurídica de los datos varía mucho entre los diferentes países. Sin embargo, la mayoría de las jurisdicciones realmente garantizan algunos derechos en los datos (cuando se trate de un conjunto).

Esta distinción entre el “contenido” de una base de datos y la colección es especialmente importante para bases de datos fácticas ya que ninguna jurisdicción garantiza un monopolio sobre los hechos individuales (“el contenido”), aunque sí podría garantizar derecho(s) sobre ellos como una colección. Para ilustrarlo, considera los simples ejemplos de una base de datos que lista los puntos de fusión de varias sustancias. Mientras que la base de datos como un todo puede estar protegida por ley para que uno no pueda acceder a ella, reutilizarla o redistribuirla sin permiso, esto no quita que uno pueda establecer que la sustancia Y se funde a la temperatura Z.

Las formas de protección pueden agruparse en dos casos:

- Derechos de autor para las compilaciones
- Un *sui generis* correcto para la colección de datos

Como hemos hecho incapié, no existen reglas generales y la situación varía según la jurisdicción. Es por esto que se procede país por país, detallando cuál (si es que existe) es el enfoque utilizado en una jurisdicción en particular.

Finalmente, debemos señalar que ante la ausencia de una protección legal, algunos proveedores de bases de datos, son capaces

de utilizar un contrato simple combinado con ciertas disposiciones legales que prohíben la violación de los mecanismos de control de acceso para lograr resultados similares a un derecho de propiedad intelectual. Por ejemplo, si X es un proveedor de una base de datos, puede alcanzar un conjunto de términos de condiciones, simplemente por:

- a) Los usuarios deben loguearse a través de una contraseña
- b) Sólo proporcionar una cuenta a un usuario con una contraseña con la condición de que el usuario acepte los términos y condiciones

Para leer más acerca de la situación jurisdiccional: “Guía para la licencia de Datos Abiertos”¹

¹Guide to Open Data Licensing: <http://opendefinition.org/guide/data/>

